

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2002, 15:48:01 ; Search time 3102.12 Seconds
(without alignments)
3130.089 Million cell updates/sec

Title: US-09-617-174b-1_copy_451_914

Perfect score: 464

Sequence: 1 ttgataatcctaatacacagg.....gcaacttcagcgtgaatcat 464

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vi:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pi:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vi:*

30: em_htg_hum:*

31: em_htg_inv:*

32: em_htg_other:*

33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description

1	462.4	99.7	170269	9	AC090307	AC090307 Homo sapi
2	462.4	99.7	173131	2	AC036176	AC036176 Homo sapi
3	462.4	99.7	173836	2	AP001404	AP001404 Homo sapi
4	390	84.1	214212	2	AC015793	AC015793 Homo sapi
5	340.8	73.4	12426	6	AX251236	AX251236 Sequence
6	313.6	67.6	12426	6	AX251235	AX251235 Sequence
7	75.2	16.2	214212	2	AC015793	AC015793 Homo sapi
8	44.4	9.6	170136	9	AC003082	AC003082 Human BAC
9	42.4	9.6	225053	2	AC084396	AC084396 Homo sapi
10	42.6	9.2	164339	2	AC025792	AC025792 Homo sapi
11	42.4	9.1	166982	9	AC084783	AC084783 Homo sapi
12	42.2	9.1	22601	8	AB022222	AB022222 Arabidops
13	41.4	8.9	7218	6	I66494	I66494 Sequence 14
14	41.4	8.9	195470	2	AC022259	AC022259 Homo sapi
15	41.2	8.9	122258	2	AL356427	AL356427 Homo sapi
16	40.6	8.8	154255	9	HS428A13	282209 Human DNA s
17	40.6	8.8	159906	2	AL590706	AL590706 Homo sapi
18	40.6	8.8	160517	2	AL589785	AL589785 Homo sapi
19	40.6	8.8	183704	2	AC023326	AC023326 Homo sapi
20	40.4	8.7	108253	9	AL136164	AL136164 Human DNA
21	40.2	8.7	89747	9	AL591022	AL591022 Human DNA
22	40	8.6	131711	2	AL445194	AL445194 Homo sapi
23	40	8.6	194515	2	AC095027	AC095027 Canis fam
24	39.8	8.6	113880	3	PFMAL3P4	AL008970 Plasmodiu
25	39.8	8.6	166125	9	AC023932	AC023932 Homo sapi
26	39.6	8.5	10127	10	AF250369S6	AF249893 Mus muscu
27	39.6	8.5	98905	2	AC095617	AC095617 Rattus no
28	39.6	8.5	110000	2	AC098517	AC098517 Rattus no
29	39.6	8.5	133751	9	HS1131C10	AL121931 Human DNA
30	39.4	8.5	74992	2	AC069574	AC069574 Homo sapi
31	39.4	8.5	154235	9	HS428A13	282209 Human DNA s
32	39.4	8.5	160517	2	AL589785	AL589785 Homo sapi
33	39.4	8.5	183704	9	AC023326	AC023326 Homo sapi
34	39.4	8.5	185676	9	AC073927	AC073927 Homo sapi
35	39.2	8.4	24023	9	AC105392	AC105392 Homo sapi
36	39.2	8.4	153003	2	AC094798	AC094798 Rattus no
37	39.2	8.4	172051	30	AC044914	AC044914 Homo sapi
38	39	8.4	26203	2	PFMAL13P7	U28971 Caenorhabdi
39	39	8.4	43547	3	U28971	U28971 Caenorhabdi
40	39	8.4	160696	2	AC107082	AC107082 Homo sapi
41	39	8.4	165046	2	AC091991	AC091991 Homo sapi
42	39	8.4	169755	2	AP001199	AP001199 Homo sapi
43	39	8.4	172915	2	AC099742	AC099742 Papio cyn
44	39	8.4	188207	2	AC021800	AC021800 Homo sapi
45	38.8	8.4	213793	9	AC019212	AC019212 Homo sapi

ALIGNMENTS

RESULT	1	AC090307	170269 bp	DNA	linear	PRI 30-DEC-2001
LOCUS	AC090307	Homo sapiens	chromosome 18,	clone RP11-851B10,	complete	sequence.
DEFINITION	AC090307	Homo sapiens	chromosome 18,	clone RP11-851B10,	complete	sequence.
ACCESSION	AC090307	HTG.				
VERSION	AC090307.7	GI:16041402				
KEYWORDS	HTG.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	1	(bases 1 to 170269)				
AUTHORS	Birren,B., Linton,L., Nusbaum,C. and Lander,E.					
TITLE	Homo sapiens chromosome 18,					
JOURNAL	Unpublished					
REFERENCE	2	(bases 1 to 170269)				
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S., Barina,N., Bastien,V., Boguslavsky,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dekrellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,					

Jones, C., Karatas, A., LaRocque, K., Lamazares, R., Landers, T.,
Lehoczy, J., Levine, R., Liu, G., MacLean, C., Macdonald, P.,
Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K.,
McPheeters, R., Melidrim, J., Meneus, L., Mihova, T., Mlenga, V.,
Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Plerre, N., Pollara, V., Raymond, C., Retta, R.,
Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M.,
Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P.,
Sougnez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A.,
Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.

TITLE
Direct Submission
JOURNAL
Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE
3 (bases 1 to 170269)
AUTHORS
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgaltier, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Choepe, Y., Collangelo, M., Collins, S., Collimore, A., Cook, A.,
Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Melidrim, J.,
Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
Direct Submission
JOURNAL
Submitted (30-DEC-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT
On Oct 11, 2001 this sequence version replaced gi:15290867.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu

Project Information
Center project name: L12576
Center clone name: 851_P_10

FEATURES
Source
1. 170269
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="18"
/clone="RP11-851B10"
/clone_lib="RPC1-11 Human Male BAC"
complement(2..177)
/rpt_family="L1MC5"
complement(189..489)
/rpt_family="AluSx"
complement(708..803)
/rpt_family="MIR"
complement(1616..1904)
/rpt_family="AluJo"
complement(2815..3120)

repeat_region
3218..3365
/rpt_family="AluSx"
repeat_region
3565..3590
/rpt_family="MIR"
repeat_region
complement(3782..4028)
/rpt_family="MIR"
repeat_region
4056..4177
/rpt_family="MIRSA"
repeat_region
4601..5620
/rpt_family="HSMAR2"
repeat_region
5621..5643
/rpt_family="CAA"
repeat_region
5644..5866
/rpt_family="HSMAR2"
repeat_region
6184..6475
/rpt_family="L1MA2"
repeat_region
6677..6837
/rpt_family="TA"
repeat_region
6852..7018
/rpt_family="TA"
repeat_region
6994..7122
/rpt_family="CATATA"
repeat_region
7146..7332
/rpt_family="CATATA"
repeat_region
7371..7526
/rpt_family="TA"
repeat_region
complement(7527..7835)
/rpt_family="AluSx"
repeat_region
7936..8080
/rpt_family="MIR"
repeat_region
8697..8927
/rpt_family="AluSx"
repeat_region
complement(9036..9288)
/rpt_family="AluSx"
repeat_region
9873..9934
/rpt_family="GA"
repeat_region
10340..10367
/rpt_family="AT-rich"
repeat_region
10368..10686
/rpt_family="AluY"
repeat_region
10687..10711
/rpt_family="AT-rich"
repeat_region
complement(10870..11043)
/rpt_family="L2"
repeat_region
11348..11375
/rpt_family="T"
repeat_region
12359..12449
/rpt_family="L3"
repeat_region
complement(12634..12890)
/rpt_family="MIR"
repeat_region
complement(13436..13725)
/rpt_family="AluJB"
repeat_region
14244..14297
/rpt_family="(CGGGG)n"
repeat_region
complement(15188..15350)
/rpt_family="MIR3"
repeat_region
16056..16114
/rpt_family="MIR3"
repeat_region
complement(16129..16429)
/rpt_family="AluSx"
repeat_region
16532..16646
/rpt_family="MER94"
repeat_region
16855..17145
/rpt_family="L2"
repeat_region
17377..17572
/rpt_family="MIR3"
repeat_region
complement(20036..20330)
/rpt_family="Charliel"
repeat_region
complement(20333..20486)
/rpt_family="Charliel"
repeat_region
20825..20691
/rpt_family="L2"


```
RESULT 5
AX251236/c
LOCUS AX251236 12426 bp DNA linear PAT 05-OCT-2001
DEFINITION Sequence 204 from Patent WO0168912.
ACCESSION AX251236
VERSION AX251236.1 GI:15984659
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct.
artificial sequence.
REFERENCE 1 (bases 1 to 12426)
Olek,A., Piepenbrock,C. and Berlin,K.
AUTHORS Diagnosis of diseases associated with tumor suppressor genes and
TITLE oncogenes
JOURNAL Patent: WO 0168912-A 204 20-SEP-2001;
EpiGenomics AG (DE)
FEATURES
Location/Qualifiers
source
1..12426
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 3791 a 137 c 2517 g 5981 t
ORIGIN
Query Match 73.4%; Score 340.8; DB 6; Length 12426;
Best Local Similarity 83.4%; Pred. No. 7.2e-81;
Matches 387; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
QY 1 ttagtaacctaatcaccagagattttaaagaagaactctctgcgcaccctaaacaata 60
Db 8056 TAAATTCCTATCAATAAATTTTAAAAAATCTCTACGCCACCCATTAACAATA 7997
QY 61 tctctaccaatttgtaataatttttgctaagtagtaacctaattttagtaggcact 120
Db 7996 TCTCTACCAATTTAATAATAATATTTTACTAATAATACCTAATTTTAAATAAACACT 7937
QY 121 gttgttatacatatccactctctcttttttgattgctcttctgttttaaggcgacta 180
Db 7936 ATATTATACATATATCCATCTCTCTTTTAAATTTATCTTCTATTATTAACAACATA 7877
QY 181 cctctcttgagcatctgcaggaatgagctgctgcagttcacacaaaagaatggagatcag 240
Db 7876 CCTCTCTTAACATCTTAACAAAATAAACTACTACAATTTTACAAAAAATAAAATCAA 7817
QY 241 agtactttttgtgcacacacgctgctcgagaaatttgtagttgtagttactatcatcacact 300
Db 7816 AATACCTTTTATMCCACCAACGATCTCAAAAATTTATAATATTACTATCATCATCACACT 7757
QY 301 acttttatttcacgaattttcaccttccggtccgctggtggccgagagattgcccgtta 360
Db 7756 ACTTTTATTCATCGAATATTTTCACCTTCGATCCTACGTAACCCGAAAAAATTACCGTA 7697
QY 361 cgcattgctgtaagtgcgttaactacacagccctctctgcgcacacattgttggagc 420
Db 7696 CGCATATCTATCGTATACATATAAATCAACACCCCTTCTACCCCAACATATAAANAAC 7637
QY 421 cttttggaagctgtgcagacacacagcaacttcaagctgaatcat 464
Db 7636 CTTTAAAAAATAATAAACAACAATAAATTTCAACCTAAATCAT 7593
RESULT 6
AX251235
LOCUS AX251235 12426 bp DNA linear PAT 05-OCT-2001
DEFINITION Sequence 203 from Patent WO0168912.
ACCESSION AX251235
VERSION AX251235.1 GI:15984658
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct.
artificial sequence.
REFERENCE 1 (bases 1 to 12426)
Olek,A., Piepenbrock,C. and Berlin,K.
AUTHORS Diagnosis of diseases associated with tumor suppressor genes and
TITLE oncogenes
JOURNAL Patent: WO 0168912-A 204 20-SEP-2001;
EpiGenomics AG (DE)
FEATURES
Location/Qualifiers
source
1..12426
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 3791 a 137 c 2517 g 5981 t
ORIGIN
Query Match 73.4%; Score 340.8; DB 6; Length 12426;
Best Local Similarity 83.4%; Pred. No. 7.2e-81;
Matches 387; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
QY 1 ttagtaacctaatcaccagagattttaaagaagaactctctgcgcaccctaaacaata 60
Db 8056 TAAATTCCTATCAATAAATTTTAAAAAATCTCTACGCCACCCATTAACAATA 7997
QY 61 tctctaccaatttgtaataatttttgctaagtagtaacctaattttagtaggcact 120
Db 7996 TCTCTACCAATTTAATAATAATATTTTACTAATAATACCTAATTTTAAATAAACACT 7937
QY 121 gttgttatacatatccactctctcttttttgattgctcttctgttttaaggcgacta 180
Db 7936 ATATTATACATATATCCATCTCTCTTTTAAATTTATCTTCTATTATTAACAACATA 7877
QY 181 cctctcttgagcatctgcaggaatgagctgctgcagttcacacaaaagaatggagatcag 240
Db 7876 CCTCTCTTAACATCTTAACAAAATAAACTACTACAATTTTACAAAAAATAAAATCAA 7817
QY 241 agtactttttgtgcacacacgctgctcgagaaatttgtagttgtagttactatcatcacact 300
Db 7816 AATACCTTTTATMCCACCAACGATCTCAAAAATTTATAATATTACTATCATCATCACACT 7757
QY 301 acttttatttcacgaattttcaccttccggtccgctggtggccgagagattgcccgtta 360
Db 7756 ACTTTTATTCATCGAATATTTTCACCTTCGATCCTACGTAACCCGAAAAAATTACCGTA 7697
QY 361 cgcattgctgtaagtgcgttaactacacagccctctctgcgcacacattgttggagc 420
Db 7696 CGCATATCTATCGTATACATATAAATCAACACCCCTTCTACCCCAACATATAAANAAC 7637
QY 421 cttttggaagctgtgcagacacacagcaacttcaagctgaatcat 464
Db 7636 CTTTAAAAAATAATAAACAACAATAAATTTCAACCTAAATCAT 7593
```

```
AUTHORS Olek,A., Piepenbrock,C. and Berlin,K.
TITLE Diagnosis of diseases associated with tumor suppressor genes and
oncogenes
JOURNAL Patent: WO 0168912-A 203 20-SEP-2001;
EpiGenomics AG (DE)
FEATURES
Location/Qualifiers
source
1..12426
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 3527 a 137 c 2591 g 6171 t
ORIGIN
Query Match 57.6%; Score 313.6; DB 6; Length 12426;
Best Local Similarity 79.7%; Pred. No. 1.4e-73;
Matches 370; Conservative 0; Mismatches 94; Indels 0; Gaps 0;
QY 1 ttagtaacctaatcaccagagattttaaagaagaactctctgcgcaccctaaacaata 60
Db 4371 TGAGTAATTTTAAATATAGGATTTTAAAAAGAAATTTTTCGCTTATTTTAAATAATA 4430
QY 61 tctctaccaatttgtagtaataatttttgctaagtagtaacctaattttagtaggcact 120
Db 4431 TTTTATTAATTTTGGTAGTAAATATTTTCTTAATAGTATTTTAAATTTTAGGTAGGTAT 4490
QY 121 gttgttatacatatccactctctcttttttgattgctcttctgttttaaggcgacta 180
Db 4491 GGTATTATATATATATTTTATTTTTCATTTGTTTTCATTTGTTTTCATTTGTTTTCAT 4550
QY 181 cctctcttgagcatctgcaggaatgagctgctgcagttcacacaaaagaatggagatcag 240
Db 4551 TTTTATTTTGGTATTTAGTAGAATGAGTCTTCTGTAGTTTATATAAAGAAATGAGATTAG 4610
QY 241 agtactttttgtgcacacacgctgctcgagaaatttgtagttgtagttactatcatcacact 300
Db 4611 AGTATTTTTCGTTGTTTATTAACCGTGTTCGAAATTTTGTAGTGTATTTATTTATATATAT 4670
QY 301 acttttatttcacgaattttcaccttccggtccgctggtggccgagagattgcccgtta 360
Db 4671 ATTTTATTTTATTCGAAATATTTTATTTTTCGCTTTTCGCTGCGGTCGAGAGGATTGTCGTA 4730
QY 361 cgcattgctgtaagtgcgttaactacacagccctctctgcgcacacattgttggagc 420
Db 4731 CGTATGCTTTGTCAGTATGATGTAATTTATAGTATTTTTCGCAATATGTTGGAGGT 4790
QY 421 cttttggaagctgtgcagacacacagcaacttcaagctgaatcat 464
Db 4791 TTTTGGAAAGTTGTGTAGATAATAGTAATTTTAGTTTGAATAT 4834
RESULT 7
AC015793 214212 bp DNA linear HTG 13-JUL-2000
LOCUS Homo sapiens chromosome 18 clone RP11-1117D15 map 18, LOW-PASS
DEFINITION SEQUENCE SAMPLING.
ACCESSION AC015793
VERSION AC015793.2 GI:7144769
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 214212)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 18, clone RP11-1117D15
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 214212)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferrelira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
```

Galagan, J., Gardyna, S., Grant, G., Hago, B., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lehoczy, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

Direct Submission
Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 3, 2000 this sequence version replaced gi:6446894.

All repeats were identified using RepeatMasker:

Smith, A. F. A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: W1BR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence-submissions@genome.wi.mit.edu

----- Project Information

Center project name: L1590

Center clone name: 1117_D_15

* NOTE: This record contains 240 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 695: contig of 695 bp in length
696 795: gap of 100 bp
796 1513: contig of 718 bp in length
1514 1613: gap of 100 bp
1614 2337: contig of 724 bp in length
2338 2437: gap of 100 bp
2438 3161: contig of 724 bp in length
3162 3261: gap of 100 bp
3262 3966: contig of 705 bp in length
3967 4066: gap of 100 bp
4067 4772: contig of 706 bp in length
4773 4872: gap of 100 bp
4873 5564: contig of 692 bp in length
5565 5664: gap of 100 bp
5665 6397: contig of 733 bp in length
6398 6497: gap of 100 bp
6498 7202: contig of 705 bp in length
7203 7302: gap of 100 bp
7303 8028: contig of 726 bp in length
8029 8128: gap of 100 bp
8129 8803: contig of 675 bp in length
8804 8903: gap of 100 bp
8904 9613: contig of 710 bp in length
9614 9713: gap of 100 bp
9714 10409: contig of 696 bp in length
10410 10509: gap of 100 bp
10510 11237: contig of 728 bp in length
11238 11337: gap of 100 bp
11338 12043: contig of 706 bp in length
12044 12143: gap of 100 bp
12144 12867: contig of 724 bp in length
12868 12967: gap of 100 bp
12968 13660: contig of 693 bp in length
13661 13760: gap of 100 bp
13761 14489: contig of 729 bp in length
14490 14589: gap of 100 bp
14590 15314: contig of 725 bp in length

15315 15414: gap of 100 bp
15415 16129: contig of 715 bp in length
16130 16229: gap of 100 bp
16230 16946: contig of 717 bp in length
16947 17046: gap of 100 bp
17047 17769: contig of 723 bp in length
17770 17869: gap of 100 bp
17870 18574: contig of 705 bp in length
18575 18674: gap of 100 bp
18675 19379: contig of 705 bp in length
19380 19479: gap of 100 bp
19480 20186: contig of 707 bp in length
20187 20286: gap of 100 bp
20287 20980: contig of 694 bp in length
20981 21080: gap of 100 bp
21081 21779: contig of 699 bp in length
21780 21879: gap of 100 bp
21880 22594: contig of 715 bp in length
22595 22694: gap of 100 bp
22695 23399: contig of 705 bp in length
23400 23499: gap of 100 bp
23500 24217: contig of 718 bp in length
24218 24317: gap of 100 bp
24318 25026: contig of 709 bp in length
25027 25126: gap of 100 bp
25127 25836: contig of 710 bp in length
25837 25936: gap of 100 bp
25937 26637: contig of 701 bp in length
26638 26737: gap of 100 bp
26738 27446: contig of 709 bp in length
27447 27546: gap of 100 bp
27547 28274: contig of 728 bp in length
28275 28374: gap of 100 bp
28375 29095: contig of 721 bp in length
29096 29195: gap of 100 bp
29196 29917: contig of 722 bp in length
29918 30017: gap of 100 bp
30018 30724: contig of 707 bp in length
30725 30824: gap of 100 bp
30825 31529: contig of 705 bp in length
31530 31629: gap of 100 bp
31630 32324: contig of 695 bp in length
32325 32424: gap of 100 bp
32425 33116: contig of 692 bp in length
33117 33216: gap of 100 bp
33217 33936: contig of 720 bp in length
33937 34036: gap of 100 bp
34037 34750: contig of 714 bp in length
34751 34850: gap of 100 bp
34851 35578: contig of 728 bp in length
35579 35678: gap of 100 bp
35679 36393: contig of 715 bp in length
36394 36493: gap of 100 bp
36494 37203: contig of 710 bp in length
37204 37303: gap of 100 bp
37304 37965: contig of 662 bp in length
37966 38065: gap of 100 bp
38066 38781: contig of 716 bp in length
38782 38881: gap of 100 bp
38882 39590: contig of 709 bp in length
39591 39690: gap of 100 bp
39691 40403: contig of 713 bp in length
40404 40503: gap of 100 bp
40504 41238: contig of 735 bp in length
41239 41338: gap of 100 bp
41339 42064: contig of 726 bp in length
42065 42164: gap of 100 bp
42165 42880: contig of 716 bp in length
42881 42980: gap of 100 bp
42981 43687: contig of 707 bp in length
43688 43787: gap of 100 bp
43788 44504: contig of 717 bp in length
44505 44604: gap of 100 bp

TITLE
JOURNAL
COMMENT

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 225053)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 225053)
Waterston,R.H.
Direct Submission
Submitted (30-OCT-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Apr 5, 2001 this sequence version replaced gi:12863223.

COMMENT

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc/index.shtml>
----- Project Information -----
Center project name: H_NH0134N03
----- Summary Statistics -----
Sequencing vector: M13; 91%
Sequencing vector: plasmid; 4%
Chemistry: Dye-terminator Big Dye; 4% of reads
Chemistry: Dye-terminator Big Dye; 4% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 184917 bases at least Q40
Consensus quality: 196549 bases at least Q30
Consensus quality: 203318 bases at least Q20
Insert size: 196000; agarose-ff
Insert size: 219053; sum-of-contigs
Quality coverage: 3.56 in Q20 bases; agarose-ff
Quality coverage: 2.91 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 61 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 4207: contig of 4207 bp in length
* 4208 4307: gap of unknown length
* 4308 10132: contig of 5825 bp in length
* 10133 10232: gap of unknown length
* 10233 15820: contig of 5588 bp in length
* 15821 15920: gap of unknown length
* 15921 20334: contig of 4414 bp in length
* 20335 20434: gap of unknown length
* 20435 27845: contig of 7411 bp in length
* 27846 27945: gap of unknown length
* 27946 32177: contig of 4232 bp in length
* 32178 32277: gap of unknown length
* 32278 38504: contig of 6227 bp in length
* 38505 38604: gap of unknown length
* 38605 47769: contig of 9165 bp in length
* 47770 47869: gap of unknown length
* 47870 55641: contig of 7772 bp in length
* 55642 55741: gap of unknown length
* 55742 65119: contig of 9378 bp in length
* 65120 65219: gap of unknown length
* 65220 73704: contig of 8485 bp in length
* 73705 73804: gap of unknown length
* 73805 83677: contig of 9873 bp in length
* 83678 83777: gap of unknown length
* 83778 92194: contig of 8417 bp in length
* 92195 92294: gap of unknown length
* 92295 100513: contig of 8219 bp in length
* 100514 100614: gap of unknown length
* 100615 106583: contig of 5970 bp in length
* 106584 106683: gap of unknown length
* 106684 117190: contig of 10507 bp in length
* 117191 117290: gap of unknown length
* 117291 117291: contig of 1438 bp in length

118729 118828: gap of unknown length
* 118829 120265: contig of 1437 bp in length
* 120266 120365: gap of unknown length
* 120366 121652: contig of 1287 bp in length
* 121653 121752: gap of unknown length
* 121753 122860: contig of 1108 bp in length
* 122861 122960: gap of unknown length
* 122961 124298: contig of 1338 bp in length
* 124299 124398: gap of unknown length
* 124399 125617: contig of 1219 bp in length
* 125618 125717: gap of unknown length
* 125719 127326: contig of 1609 bp in length
* 127327 127426: gap of unknown length
* 127427 129643: contig of 2217 bp in length
* 129644 129743: gap of unknown length
* 129744 130956: contig of 1213 bp in length
* 130957 131056: gap of unknown length
* 131057 132284: contig of 1228 bp in length
* 132285 132384: gap of unknown length
* 132385 134025: contig of 1641 bp in length
* 134026 134125: gap of unknown length
* 134126 135771: contig of 1646 bp in length
* 135772 135871: gap of unknown length
* 135872 137168: contig of 1297 bp in length
* 137169 137268: gap of unknown length
* 137269 138776: contig of 1508 bp in length
* 138777 138876: gap of unknown length
* 138877 140630: contig of 1754 bp in length
* 140631 140730: gap of unknown length
* 140731 143390: contig of 2660 bp in length
* 143391 143490: gap of unknown length
* 143491 144885: contig of 1395 bp in length
* 144886 144985: gap of unknown length
* 144986 146277: contig of 1292 bp in length
* 146278 146377: gap of unknown length
* 146378 148790: contig of 2413 bp in length
* 148791 148890: gap of unknown length
* 148891 151618: contig of 2728 bp in length
* 151619 151718: gap of unknown length
* 151719 153528: contig of 1810 bp in length
* 153529 153628: gap of unknown length
* 153629 155632: contig of 2004 bp in length
* 155633 155732: gap of unknown length
* 155733 158353: contig of 2621 bp in length
* 158354 158453: gap of unknown length
* 158454 161521: contig of 3068 bp in length
* 161522 161621: gap of unknown length
* 161622 163929: contig of 2308 bp in length
* 163930 164029: gap of unknown length
* 164030 165740: contig of 1611 bp in length
* 165641 165740: gap of unknown length
* 165741 168599: contig of 2959 bp in length
* 168600 168799: gap of unknown length
* 168800 169971: contig of 1172 bp in length
* 169972 170071: gap of unknown length
* 170072 171383: contig of 1312 bp in length
* 171384 171483: gap of unknown length
* 171484 174613: contig of 3130 bp in length
* 174614 174713: gap of unknown length
* 174714 178023: contig of 3310 bp in length
* 178024 178123: gap of unknown length
* 178124 179809: contig of 1686 bp in length
* 179810 179909: gap of unknown length
* 179910 182166: contig of 2257 bp in length
* 182167 182266: gap of unknown length
* 182267 184778: contig of 2512 bp in length
* 184779 184878: gap of unknown length
* 184879 187289: contig of 2411 bp in length
* 187290 187389: gap of unknown length
* 187390 190025: contig of 2636 bp in length
* 190026 190125: gap of unknown length
* 190126 193628: contig of 3503 bp in length
* 193629 193728: gap of unknown length

* 193729 196385: contig of 2657 bp in length
 * 196386 196485: gap of unknown length
 * 196486 199287: contig of 2802 bp in length
 * 199288 199387: gap of unknown length
 * 199388 202073: contig of 2686 bp in length
 * 202074 202173: gap of unknown length
 * 202174 205921: contig of 3748 bp in length
 * 205922 206021: gap of unknown length
 * 206022 211281: contig of 5260 bp in length
 * 211282 211381: gap of unknown length
 * 211382 215785: contig of 4404 bp in length
 * 215786 215885: gap of unknown length
 * 215886 220968: contig of 5083 bp in length
 * 220969 221068: gap of unknown length
 * 221069 225053: contig of 3985 bp in length.

FEATURES

source

1. .225053
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosomes="UNK"
 /clone="RP11-134N3"
 1. .4207
 /note="assembly_name:Contig100"
 clone_end:r7
 vector_side:left
 4308. .10132
 /note="assembly_name:Contig101"
 10233. .15820
 /note="assembly_name:Contig102"
 15921. .20334
 /note="assembly_name:Contig103"
 20435. .27845
 /note="assembly_name:Contig104"
 27946. .32177
 /note="assembly_name:Contig105"
 32278. .38504
 /note="assembly_name:Contig106"
 38605. .47769
 /note="assembly_name:Contig107"
 47870. .55641
 /note="assembly_name:Contig108"
 55742. .65119
 /note="assembly_name:Contig109"
 65220. .73704

Query Match 9.6%; Score 44.4; DB 2; Length 225053;
 Best Local Similarity 46.7%; Pred. No. 0.26;
 Matches 141; Conservative 0; Mismatches 161; Indels 0; Gaps 0;
 QY 1 t9agtaatcctaatacagagattttaaaagaactctgcgcacccattaaacaata 60
 DB 194153 TAAGTAATCTCAGTACATATTTCTAGAAAAAGAGATACAAATATTTTCTCTTGC 194094
 QY 61 tctctaccaatttggtagtaataattttgctaatagtacacctaatttttagtgtaggcact 120
 DB 194093 ACTCTTATAGTGCTTCTCAAAATATTTTTTAAAGTCATCTTAGTTTTTATTTCTTATA 194034
 QY 121 ggtttatacatatcatcattctcttttttgattgcttcttctgttaatgggcagcta 180
 DB 194033 TTCACAATGCTTAAATAGATTTTGTAGACTCTCATGCTGAATGATGTGTAGATT 193974
 QY 181 cctctctggcatctagcagaatgagctgcgagtttacacaaaagaagatgagatcag 240
 DB 193973 CAACTGTGCTAGTTTAAACACACAGCTATGCTCATCTCTTTGTGAAGCTTAAATATGTG 193914
 QY 241 agtactttttgtgccacaacgctgtctgagaaatttgtagttactatcatcacatt 300
 DB 193913 TCTGTCTTTTATCCAAATAGTCCCTGTTAACTCAGTCTGTGTAACAATTTATCAGATT 193854
 QY 301 ac 302
 DB 193853 CC 193852

RESULT 10

AC025792/c

LOCUS

DEFINITION

AC025792

AC025792

AC025792.2

HTG: HTGS_PHASE1; HTGS_DRAFT.

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC025792 164339 bp DNA linear HTG 06-APR-2000
 Homo sapiens chromosome 11 clone RP11-243M7 map 11, WORKING DRAFT
 SEQUENCE, 13 unordered pieces.
 AC025792
 HTG: HTGS_PHASE1; HTGS_DRAFT.
 AC025792.2 GI:7467488
 human.

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 164339)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Homo sapiens chromosome 11, clone RP11-243M7

Unpublished
 2 (bases 1 to 164339)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
 Boguslavskiy,L., Boukhgaltier,B., Brown,A., Burkett,G.,
 Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
 Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
 Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
 Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
 Meidrim,J., Menes,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Plerre,N.,
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE

JOURNAL

COMMENT

Submitted (14-MAR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Apr 6, 2000 this sequence version replaced gi:7239678.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR

Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu

Center project name: L7901
 Center clone name: 243_M7

Sequencing vector: M13; M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731

Consensus quality: 156382 bases at least Q40
 Consensus quality: 160561 bases at least Q30
 Consensus quality: 162162 bases at least Q20

Insert size: 170000; agarose-fp
 Quality coverage: 4.6 in Q20 bases; agarose-fp
 Quality coverage: 4.7 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 13 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will

RESULT	11
AC084783	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	

ORGANISM homo sapiens

REFERENCE
1 (bases 1 to 166982)
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
AUTHORS
Rowen, L., Madan, A., Qin, S., Baradaran, L., Birditt, B., Bloom, S.,
Dors, M., Dickhoff, R., Fleetwood, P., Harrison, G., Kaur, A., Madan, A.,
Nesbitt, R., Traicoff, R. and Hood, L.
TITLE
Sequencing of human chromosome 15 D15S146-D15S117 region

JOURNAL UNPUBLISHED
REFERENCE 2. (bases 1 to 166982)

2 pages (1 to 2 pages)
 REFERENCE
 AUTHORS
 ROWEN, L., MADAN, A., QIN, S., BARADERANI, L., BIRDITT, B., BLOOM, S.,
 DORS, M., DICKHOFF, R., FLEETWOOD, P., HARRISON, G., KAUF, A., MADAN, A.,
 NESBITT, R., TRALCOFF, R. and HOOD, L.
 Direct Submission
 Submitted (16-NOV-2000) Multimegabase Sequencing Center, Institute
 for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
 98105, USA

JOURNAL Submitted (13-JAN-2001) Multimegabase Sequencing Center, Institut
for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, W
98105, USA
COMMENT On Jan 13, 2001 this sequence version replaced qi.11811781.

```

----- Summary Statistics -----
Sequencing vector: pUC18; L08752
Chemistry: dye-terminator Big Dye; 90% of reads
Chemistry: Dye-primer Big Dye; 10% of reads
Assembly program: Phrap; version 0.990399

```

.....
This clone was finished using overlapping sequence from AC036163
[drafting center=WIBR] and AC068726 [drafting center = UWMSC]

FEATURES	Location/Qualifiers
source	1..166982

```

source
1. 100502
/organism="Homo sapiens"
/db xref="taxon:9606"

```

```
/map="15q21.3"
/clone="RPCI human BAC library 11"
```


DEFINITION	Homo sapiens chromosome 2 clone RP11-335E2 map 2, *** SEQUENCING IN
ACCESSION	PROGRESS ***, 58 unordered pieces.
VERSION	AC022259
KEYWORDS	AC022259.1 GI:6778490
SOURCE	HTG: HTGS_PHASE1.
ORGANISM	human.
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE	1 (bases 1 to 195470)
AUTHORS	Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE	Homo sapiens chromosome 2, clone RP11-335E2
REFERENCE	2 (bases 1 to 195470)
AUTHORS	Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
	Anderson, S., Baldwin, J., Barna, N., Becker, R., Beda, F.,
	Boguslavsky, L., Bouckgalter, B., Brown, A., Burkett, G., Castle, A.,
	Choepel, Y., Collangelo, M., Collins, S., Collymore, A., Cooke, P.,
	DeArelano, K., Dewar, K., Domino, M., Doyle, M., Fenestor, J.,
	Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J.,
	Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
	Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
	Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K.,
	Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,
	McPheeters, R., Meldrim, J., Meneus, L., Morrow, J., Naylor, J.,
	Norman, C. H., O'Connor, T., O'Donnell, P., Olivar, T. M., Peterson, K.,
	Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
	Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
	Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
	Tjrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,
	Zimmer, A. and Zody, M.
TITLE	Direct Submission
JOURNAL	Submitted (27-JAN-2000) Whitehead Institute/MIT Center for Genome
COMMENT	Research, 320 Charles Street, Cambridge, MA 02141, USA
	All repeats were identified using RepeatMasker:
	Smith, A. F. A. & Green, P. (1996-1997)
	http://ftp.genome.washington.edu/RM/RepeatMasker.html
	----- Genome Center
	Center: Whitehead Institute/ MIT Center for Genome Research
	Center code: WIBR
	Web site: http://www-seq.wi.mit.edu
	Contact: sequence_submissions@genome.wi.mit.edu
	----- Project Information
	Center project name: L5217
	Center clone name: 335_E_2

	* NOTE: This is a 'working draft' sequence. It currently
	* consists of 58 contigs. The true order of the pieces
	* is not known and their order in this sequence record is
	* arbitrary. Gaps between the contigs are represented as
	* runs of N, but the exact sizes of the gaps are unknown.
	* This record will be updated with the finished sequence
	* as soon as it is available and the accession number will
	* be preserved.
	* 1 1169: contig of 1169 bp in length
	* gap of unknown length
	* 1170 2744: contig of 1575 bp in length
	* gap of unknown length
	* 2745 3891: contig of 1147 bp in length
	* gap of unknown length
	* 3892 4958: contig of 1067 bp in length
	* gap of unknown length
	* 4959 6042: contig of 1084 bp in length
	* gap of unknown length
	* 6043 7112: contig of 1070 bp in length
	* gap of unknown length
	* 7113 8242: contig of 1130 bp in length
	* gap of unknown length
	* 8243 9440: contig of 1198 bp in length
	* gap of unknown length
	* 9441 10503: contig of 1063 bp in length
	* gap of unknown length
	* 10504 11766: contig of 1263 bp in length
	* gap of unknown length
	* 11767 gap of unknown length
	* 12798: contig of 1032 bp in length
	* gap of unknown length
	* 14299: contig of 1501 bp in length
	* gap of unknown length
	* 15314: contig of 1015 bp in length
	* gap of unknown length
	* 16978: contig of 1664 bp in length
	* gap of unknown length
	* 18049: contig of 1071 bp in length
	* gap of unknown length
	* 19597: contig of 1548 bp in length
	* gap of unknown length
	* 21003: contig of 1406 bp in length
	* gap of unknown length
	* 22126: contig of 1123 bp in length
	* gap of unknown length
	* 23502: contig of 1376 bp in length
	* gap of unknown length
	* 24450: contig of 948 bp in length
	* gap of unknown length
	* 25766: contig of 1316 bp in length
	* gap of unknown length
	* 27235: contig of 1469 bp in length
	* gap of unknown length
	* 29226: contig of 1991 bp in length
	* gap of unknown length
	* 30961: contig of 1735 bp in length
	* gap of unknown length
	* 32608: contig of 1647 bp in length
	* gap of unknown length
	* 34331: contig of 1723 bp in length
	* gap of unknown length
	* 36533: contig of 2202 bp in length
	* gap of unknown length
	* 39299: contig of 2766 bp in length
	* gap of unknown length
	* 41451: contig of 2152 bp in length
	* gap of unknown length
	* 43295: contig of 1844 bp in length
	* gap of unknown length
	* 45678: contig of 2383 bp in length
	* gap of unknown length
	* 48551: contig of 2873 bp in length
	* gap of unknown length
	* 51186: contig of 2635 bp in length
	* gap of unknown length
	* 53207: contig of 2021 bp in length
	* gap of unknown length
	* 55697: contig of 2490 bp in length
	* gap of unknown length
	* 58328: contig of 2631 bp in length
	* gap of unknown length
	* 59424: contig of 1096 bp in length
	* gap of unknown length
	* 62209: contig of 2785 bp in length
	* gap of unknown length
	* 64795: contig of 2586 bp in length
	* gap of unknown length
	* 67972: contig of 3177 bp in length
	* gap of unknown length
	* 71707: contig of 3735 bp in length
	* gap of unknown length
	* 74928: contig of 3221 bp in length
	* gap of unknown length
	* 78667: contig of 3739 bp in length
	* gap of unknown length
	* 82959: contig of 4292 bp in length
	* gap of unknown length
	* 88848: contig of 5889 bp in length
	* gap of unknown length
	* 93807: contig of 4959 bp in length
	* gap of unknown length


```
source
1. .122258
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP11-78B14"
/clone_lib="RPC1-11.1"
1. .2843
/note="assembly_fragment:00918
fragment_chain:1"
2944. .6529
/note="assembly_fragment:00793
fragment_chain:1"
6630. .16752
/note="assembly_fragment:00592
fragment_chain:1"
16853. .19610
/note="assembly_fragment:00690
fragment_chain:1"
19711. .30027
/note="assembly_fragment:00317
fragment_chain:2"
30128. .35982
/note="assembly_fragment:00020
fragment_chain:2"
36083. .39034
/note="assembly_fragment:00521
fragment_chain:2"
39135. .41364
/note="assembly_fragment:00105
fragment_chain:3"
41465. .45036
/note="assembly_fragment:00723
fragment_chain:3"
45137. .54043
/note="assembly_fragment:00499
fragment_chain:4"
54144. .67915
/note="assembly_fragment:00122
fragment_chain:4"
68016. .71519
/note="assembly_fragment:00673
fragment_chain:5"
71620. .81674
/note="assembly_fragment:00183
fragment_chain:5"
81775. .85041
/note="assembly_fragment:00848
fragment_chain:6"
85142. .88867
/note="assembly_fragment:00370
fragment_chain:6"
88968. .94857
/note="assembly_fragment:00286"
94958. .98108
/note="assembly_fragment:00287"
98209. .100727
/note="assembly_fragment:00352"
100828. .104494
/note="assembly_fragment:00571"
104595. .107100
/note="assembly_fragment:00818"
107201. .110041
/note="assembly_fragment:00860"
110142. .114180
/note="assembly_fragment:00929"
114281. .116760
/note="assembly_fragment:00932"
116861. .122258
/note="assembly_fragment:01078"
37169 a 23367 c 22456 g 36922 t 2344 others
BASE COUNT
ORIGIN
```

```
Query Match      8.9%; Score 41.2; DB 2; Length 122258;
Best Local Similarity 47.6%; Pred. No. 1.8;
Matches 121; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

Qy 61 tctctaccaatttggtagtaataatttgcctaataatgtagtaccttaatttttaggtaggcact 120
Db 4356 TCACCTAACAGTTTGGCTATTTTAAATTTTATTCGATTTTAAACCTTTGAAGATTTTCTG 4257

Qy 121 gtgtttatacatatccattcccttcttttttgattgattgctcttcttcttgaatgggcagcta 180
Db 4296 GAATGTGATTATGCTTTCGTGTAGTCTCTTCATTTTCTTCTCTGTTTAACTGCCATGGC 4237

Qy 181 cctctcttggtcatctagcagaatgagctgctgcagttttacacaaaaagaatggagatcag 240
Db 4236 ACAGAAATTTTCAGATATTGGATCTAACAAAAAAGTATGCTATAATAACAAACAGGTAAA 4177

Qy 241 agtactttttgtgccaaccaacgtgctctgagaaatttgtagtcttactatcatcacacatt 300
Db 4176 ATACATTTTGGGATTACATTAATAAAAAAGAACTATAGGCATTTCCAGACTCTGAAGTT 4117

Qy 301 actttttatttcac 314
Db 4116 ATGATAATTAATTC 4103

Search completed: September 24, 2002, 15:52:52
Job time: 20267 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2002, 15:46:36 ; Search time 403.26 Seconds
(without alignments)
1975.520 Million cell updates/sec

Title: US-09-617-174B-1_COPY_451_914

Perfect score: 464

Sequence: 1 tggataatcctaatcacagg.....gcaacttcagcctgaatcat 464

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	464	100.0	1141	18	Human maspin promo
2	340.8	73.4	12426	22	Tumour suppressor
3	313.6	67.6	12426	22	Tumour suppressor
4	38.2	8.2	1239	22	Human immune/haema
5	38.2	8.2	13537	22	Human immune/haema
6	38	8.2	56153	22	Tumour suppressor
7	37.4	8.1	6063	22	Tumour suppressor
8	37	8.0	7143	21	Plasmodium falciparum
9	36.8	7.9	513445	22	Soybean 318013 reg

10	36.6	7.9	8964	24	ABL333534	Human immune syste
11	36.6	7.9	8964	24	AAS61276	Human gene regulat
12	36.4	7.8	44211	22	AAK85974	Human immune/haema
13	36.4	7.8	61020	22	AAS46787	Tumour suppressor
14	36.2	7.8	2234	17	AAT09022	Arabidopsis thalia
15	36.2	7.8	6309	24	ABL32304	Human immune syste
16	35.8	7.7	6127	24	ABL33615	Human immune syste
17	35.8	7.7	8123	24	ABL34227	Human immune syste
18	35.6	7.7	952	19	AAX14118	H. pylori GHPO 122
19	35.6	7.7	9293	22	AAS45320	Chemically pretrea
20	35.4	7.6	7403	22	AAS46804	Tumour suppressor
21	35.4	7.6	7403	22	ABL34217	Human immune syste
22	35.4	7.6	56153	22	AAS46794	Tumour suppressor
23	35.2	7.6	11047	22	AAS45480	Chemically pretrea
24	35.2	7.6	11047	24	ABL33985	Human immune syste
25	35.2	7.6	28091	22	ABA15375	Human nervous syst
26	35.2	7.6	28091	22	AAK77432	Human immune/haema
27	35.2	7.6	28120	22	ABA15379	Human immune/haema
28	35.2	7.6	28120	22	AAK77434	Human genome from
29	35.2	7.6	114793	22	AAK08215	Tumour suppressor
30	35	7.5	15592	22	AAS46454	Human immune syste
31	35	7.5	15592	24	ABL33327	Human immune syste
32	35	7.5	16509	24	ABL33321	Enterococcus faeca
33	34.8	7.5	1494	20	AAX20170	Human immune syste
34	34.8	7.5	6134	24	ABL33183	Human immune syste
35	34.8	7.5	6259	20	AAX13066	Enterococcus faeca
36	34.8	7.5	7069	22	AAS46654	Tumour suppressor
37	34.8	7.5	7069	24	ABL33353	Human immune syste
38	34.8	7.5	7069	24	AAS61220	Human gene regulat
39	34.6	7.5	370	22	AAI84576	Human polynucleoti
40	34.6	7.5	890	22	AAL15375	Human breast cance
41	34.6	7.5	6136	24	AAS61297	Human gene regulat
42	34.6	7.5	18028	23	ABL13284	Drosophila melanog
43	34.4	7.4	834	22	AAH03576	Human cDNA clone (
44	34.4	7.4	5947	22	AAS46676	Tumour suppressor
45	34.4	7.4	6127	24	ABL33614	Human immune syste

ALIGNMENTS

RESULT 1

AAT89000

ID AAT89000 standard; cDNA; 1141 BP.

XX AC AAT89000;

XX DT 22-APR-1998 (first entry)

XX DE Human maspin promoter and partial coding cDNA.

XX KW Maspin; serpin; mammary epithelial cell; human; promoter; malignant;

XX KW tumour cell; treatment; prostate cancer; breast cancer; metastasis; ss.

XX OS Homo sapiens.

XX FH Key

XX FT promoter

XX FT Location/Qualifiers

XX FT 1..1141

XX FT /tag= a

XX FT /note= "maspin promoter region"

XX FT misc_signal

XX FT /tag= b

XX FT /note= "Putative regulatory element AP2"

XX FT misc_signal

XX FT /tag= c

XX FT /note= "Putative regulatory element API"

XX FT misc_signal

XX FT /tag= d

XX FT /note= "Putative regulatory element Ets"

XX FT misc_signal

XX FT /tag= e

XX FT /note= "Putative regulatory element HRE"

XX FT misc_signal

XX FT /tag= e

XX FT /note= "Putative regulatory element HRE"

[illegible]

PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251899.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis.
XX
XX
XX Disclosure: SEQ ID NO 23775; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX
XX Sequence 1239 BP; 391 A; 191 C; 176 G; 481 T; 0 other;
SQ

Query Match 8.2%; Score 38.2; DB 22; Length 1239;
Best Local Similarity 48.4%; Pred. NO. 0.61;
Matches 106; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

Qy 1 tgaagtaatcctaatcacaggattttaaagaaacattcttcgcgccaccattaaacaata 60
Db 82 tcaagttttatcattgaagacatttgaagatattcttttattcaagcttaaatccat 141

Qy 61 tctctaccatttggtagtaataattttgctaaatagtaacacatttttagtaggcact 120
Db 142 ttcttaattaaattagacaacatttctttacaagaaattgaataatctgatagtact 201

Qy 121 ggtttatcacatatccattctcttttttttgattgtctttcttttaattgagcagcta 180
Db 202 ttcatagaatacaaatgttagcctttttttttgttttttttaaggaattggcagctg 261

Qy 181 cctctcttggaatctagcagaatgagctgctcagttta 219
Db 262 ttgtgagtgatcaagaagacgagtaagagagtaattta 300

RESULT 5
AAK68964
ID AAK68964 standard; DNA; 13537 BP.
XX
AC AAK68964;

XX 06-NOV-2001 (first entry)
DT Human immune/hematopoietic antigen genomic sequence SEQ ID NO:23776.
DE
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
XX WO200157182-A2.
XX 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0236279.
PR 22-AUG-2000; 2000US-0236681.
PR 22-AUG-2000; 2000US-0236868.
PR 22-AUG-2000; 2000US-0237182.
PR 23-AUG-2000; 2000US-0237009.
PR 30-AUG-2000; 2000US-0238924.
PR 01-SEP-2000; 2000US-0239287.
PR 01-SEP-2000; 2000US-0239343.
PR 01-SEP-2000; 2000US-0239344.
PR 05-SEP-2000; 2000US-0239345.
PR 05-SEP-2000; 2000US-0239509.
PR 05-SEP-2000; 2000US-0239513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.

```
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-024617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249267.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.

PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis.
XX
XX Disclosure; SEQ ID NO 23776; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
XX amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patient's own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting
XX the nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/haematopoietic-related diseases, especially
XX cancers and cancer metastases of haematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/haematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAK82169
XX represent sequences used in the exemplification of the present invention.
XX
XX Sequence 13537 BP; 4245 A; 2216 C; 2328 G; 4748 T; 0 other;
XX
XX Query Match 8.2%; Score 38.2; DB 22; Length 13537;
XX Best Local Similarity 48.4%; Pred. NO. 1.4; Indels 0; Gaps 0;
XX Matches 106; Conservative 0; Mismatches 113;
QY 1 tgaagtaactcctaacagagattttaaaagaaactctctgcgcaccattaaacaata 60
DB 3161 tagttttatcatgaagacattgtttaagtatatctttattcaaaagcttaaatccat 3220
QY 61 tctcctaccaatttggtagtaaaattttgctaaatagtaacctaatttttagtaggcact 120
DB 3221 ttcttaattaaattagacaacactttcttacaagaaaattgaataatctgatagct 3280
QY 121 gtgtttatcatatccattctctctttttttgattgtcttctgtttaatgggcagctta 180
DB 3281 ttcatagaatacaaatgtagccctttattgtttgtttttttaagcattggcagctg 3340
QY 181 cctctcttgccatctagcagaatgagctgctgagttta 219
DB 3341 ttgtgaagtcataaagaacgcgagtaaaaggagtaattta 3379
XX
XX RESULT 6
XX AAS46793
XX ID AAS46793 standard; DNA; 56153 BP.
XX
XX AC AAS46793;
XX
XX 18-DEC-2001 (first entry)
XX Tumour suppressor gene derived chemically modified sequence #519.
XX
XX Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
XX cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
```

KW cytosine methylation; ds.
 XX Homo sapiens.
 OS
 PN WO200168912-A2.
 XX
 XX 20-SEP-2001.
 XX
 XX 15-MAR-2001; 2001WO-EP02955.
 XX
 XX 15-MAR-2000; 2000DE-1013847.
 PR 06-APR-2000; 2000DE-1019058.
 PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 XX (EPIG-) EPIGENOMICS AG.
 XX
 XX Olek A, Piepenbrock C, Berlin K;
 XX
 XX WPI; 2001-602752/68.
 DR
 XX
 XX Fragments of chemically modified genes associated with tumour suppressor
 PT genes and oncogenes, useful in designing primers and probes for
 PT analysing diseases associated with cytosine methylation state e.g.
 PT cancer
 XX
 XX Claim 1; SEQ ID No 519; 27pp; English.
 PS
 XX The invention relates to a nucleic acid comprising a sequence of 18
 CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
 CC bisulphite, of genes associated with tumour suppression and
 CC oncogenes having a sequence taken from 536 (actually 533 since
 CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
 CC (Ss) and sequences complementary to (Ss). The nucleic acid may be a
 CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
 CC form part of a set of probes for detecting the cytosine methylation state
 CC and/or single nucleotide polymorphisms and also to be used in an
 CC array for analysing diseases associated with CpG dinucleotides e.g.
 CC cancers and tumours. The probes can also be used in a method for
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis
 CC and/or therapy of existing diseases or the predisposition to specific
 CC diseases, by analysing cytosine methylations. The parameters may be
 CC compared to another set of genetic and/or epigenetic parameters, the
 CC differences serving as basis for diagnosis and/or prognosis events which
 CC are disadvantageous to patients. The present sequence is one of the
 CC 533 genomic sequences derived from tumour suppressor genes and
 CC oncogenes.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 56153 BP; 13943 A; 1002 C; 14095 G; 27113 T; 0 other;
 SQ
 Query Match 8.2%; Score 38; DB 22; Length 56153;
 Best Local Similarity 47.5%; Pred. No. 2.8;
 Matches 113; Conservative 0; Mismatches 125; Indels 0; Gaps 0;
 QY 73 ttggttagtaaaatttttgcataatgactacctaatttttaggtaggcactgtgtttacacat 132
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 31726 ttgggattataggtatgagtattgttttttagttttgtttgttttttttttttttttt 31785
 QY 133 atatcattcctcttt 192
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 31786 ttattgagtt 31845
 QY 193 tctagcagaatagctgctgcagttttacacaaaagaatggagatcagagttactttttgt 252
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 Db 31846 ttctcgtagtataatttttttaattgttttagaaatgaatttaattgaattataaatttt 31905
 QY 253 gccaccaagctgctgagaaattgttagtttactatcatcacattacttttttttttt 310

Db 31906 ttaggtt 31963
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 RESULT 7
 AAS46337
 ID AAS46337 standard; DNA; 6063 BP.
 XX
 XX AAS46337;
 AC
 XX 18-DEC-2001 (first entry)
 DT
 XX Tumour suppressor gene derived chemically modified sequence #59.
 DE
 XX Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
 KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
 KW cytosine methylation; ds.
 XX
 XX Homo sapiens.
 OS
 XX WO200168912-A2.
 PN
 XX 20-SEP-2001.
 PD
 XX 15-MAR-2001; 2001WO-EP02955.
 XX
 XX 15-MAR-2000; 2000DE-1013847.
 PR 06-APR-2000; 2000DE-1019058.
 PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 XX (EPIG-) EPIGENOMICS AG.
 XX
 XX Olek A, Piepenbrock C, Berlin K;
 PI
 XX WPI; 2001-602752/68.
 DR
 XX
 XX Fragments of chemically modified genes associated with tumour suppressor
 PT genes and oncogenes, useful in designing primers and probes for
 PT analysing diseases associated with cytosine methylation state e.g.
 PT cancer
 XX
 XX Claim 1; SEQ ID No 59; 27pp; English.
 PS
 XX The invention relates to a nucleic acid comprising a sequence of 18
 CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
 CC bisulphite, of genes associated with tumour suppression and
 CC oncogenes having a sequence taken from 536 (actually 533 since
 CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
 CC (Ss) and sequences complementary to (Ss). The nucleic acid may be a
 CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
 CC form part of a set of probes for detecting the cytosine methylation state
 CC and/or single nucleotide polymorphisms and also to be used in an
 CC array for analysing diseases associated with CpG dinucleotides e.g.
 CC cancers and tumours. The probes can also be used in a method for
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis
 CC and/or therapy of existing diseases or the predisposition to specific
 CC diseases, by analysing cytosine methylations. The parameters may be
 CC compared to another set of genetic and/or epigenetic parameters, the
 CC differences serving as basis for diagnosis and/or prognosis events which
 CC are disadvantageous to patients. The present sequence is one of the
 CC 533 genomic sequences derived from tumour suppressor genes and
 CC oncogenes.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 6063 BP; 1704 A; 41 C; 1173 G; 3145 T; 0 other;
 SQ
 Query Match 8.1%; Score 37.4; DB 22; Length 6063;

PT cytosine methylation -

New nucleic acid sequences from chemically modified genes associated with gene regulation, useful for analysing cytosine methylations for diagnosis and therapy of diseases e.g. severe combined immunodeficiency disease -

XX Disclosure; SEQ ID No 237; 26pp; English.

XX The invention relates to 224 nucleic acid sequences comprising at least

XX 18 bases of a chemically pretreated gene associated with gene regulation

CC selected from 43 known genes (or complementary sequences). The

CC chemical pretreatment converts cytosine bases unmethylated at the

CC 5-position to uracil or another base with hybridisation behaviour

CC dissimilar to cytosine, to enable analysis of cytosine methylations.

CC The DNA sequences, oligomers (or sets/arrays) and method are

CC useful in the diagnosis of diseases (or predisposition to diseases)

CC associated with gene regulation and in therapy of such diseases, by

CC enabling analysis of the cytosine methylation patterns of such genes,

CC kits are provided. They are especially useful in diagnosis

CC and therapy of e.g. severe combined immunodeficiency disease, cardiac

CC disorders, haemophilia, solid tumours and cancer, Werner syndrome,

CC asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,

CC preclampsia, graft versus-host disease. The present sequence is a

CC sequence included in the sequence data for this specification and is

CC associated with the human gene regulation-associated genes.

CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic

CC format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 8964 BP; 2596 A; 89 C; 1815 G; 4464 T; 0 other;

XX SQ

Query Match 7.9%; Score 36.6; DB 24; Length 8964;

Best Local Similarity 53.1%; Pred. No. 3.6;

Matches 78; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

Qy 72 ttgtgtagtaattttgtcctaagtacctaatttttagtaggcactgtgtttatca 131

Db 7691 ttaattagtaattttgttgaattatagacggttttagggagggttttttagattga 7740

Qy 132 tatatccattcctctttttttgattgtcttttctttaaaggcagctactctcttggc 191

Db 7741 ttttttggtagattgttatttttttaggggtttttgtttattttttatttat 7800

Qy 192 atctagcagaatgagctgcagttt 218

Db 7801 ttgtaggttgtagtcggttttagttt 7827

RESULT 12

AAK85974/c

ID AAK85974 standard; DNA; 44211 BP.

XX AAK85974;

AC AAK85974;

XX 07-NOV-2001 (first entry)

XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:40786.

DE Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

XX cytostatic; gene therapy; vaccine; metastasis; ds.

XX Homo sapiens.

XX WO200157182-A2.

XX 09-AUG-2001.

XX 17-JAN-2001; 2001WO-US01354.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226868.

PR 22-AUG-2000; 2000US-0227182.

PR 23-AUG-2000; 2000US-0227009.

PR 30-AUG-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.

PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229345.

PR 05-SEP-2000; 2000US-0229509.

PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.

PR 08-SEP-2000; 2000US-0231242.

PR 08-SEP-2000; 2000US-0231243.

PR 08-SEP-2000; 2000US-0231244.

PR 08-SEP-2000; 2000US-0231413.

PR 08-SEP-2000; 2000US-0231414.

PR 08-SEP-2000; 2000US-0232080.

PR 08-SEP-2000; 2000US-0232081.

PR 12-SEP-2000; 2000US-0231968.

PR 14-SEP-2000; 2000US-0232397.

PR 14-SEP-2000; 2000US-0232398.

PR 14-SEP-2000; 2000US-0232399.

PR 14-SEP-2000; 2000US-0232400.

PR 14-SEP-2000; 2000US-0232401.

PR 14-SEP-2000; 2000US-0233063.

PR 14-SEP-2000; 2000US-0233064.

PR 14-SEP-2000; 2000US-0233065.

PR 21-SEP-2000; 2000US-0234223.

PR 21-SEP-2000; 2000US-0234274.

PR 25-SEP-2000; 2000US-0234997.

PR 25-SEP-2000; 2000US-0234998.

PR 26-SEP-2000; 2000US-0235484.

PR 27-SEP-2000; 2000US-0235834.

PR 27-SEP-2000; 2000US-0235836.

PR 29-SEP-2000; 2000US-0236327.

PR 29-SEP-2000; 2000US-0236367.

PR 29-SEP-2000; 2000US-0236368.

PR 29-SEP-2000; 2000US-0236369.

PR 29-SEP-2000; 2000US-0236370.

PR 02-OCT-2000; 2000US-0236802.

PR 02-OCT-2000; 2000US-0237037.

PR 02-OCT-2000; 2000US-0237038.

PR 02-OCT-2000; 2000US-0237039.

PR 02-OCT-2000; 2000US-0237040.

PR 13-OCT-2000; 2000US-0239935.

PR 13-OCT-2000; 2000US-0239937.

PR 20-OCT-2000; 2000US-0240960.

Wed Sep 25 08:14:17 2002

Sequence 6309 BP; 1758 A; 69 C; 1364 G; 3118 T; 0 other;

Db 4699 tttttttttttt: 4711

Search completed: September 24, 2002, 15:48:26
Job time: 15301 sec

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2002, 14:45:10 ; Search time 3309.28 Seconds
(without alignments)
1892.432 Million cell updates/sec

Title: US-09-617-174B-1_COPY_451_914

Perfect score: 464

Sequence: 1 tsgatcctcaatcacagg.....gcaacttcagcctgaatcat 464

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_htc:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_htc:*
- 12: gb_gss:*
- 13: em_gss_hum:*
- 14: em_gss_inv:*
- 15: em_gss_pln:*
- 16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43.4	9.4	892	A2539346	A2539346 ENTGP83TR
2	41.8	9.0	1201	CNS016FX	AL106695 Drosophil
3	40.6	8.8	681	A0421326	AQ421326 RPCI-II-1
4	40.2	8.7	1042	CNS024NB	AL180992 Tetraodon
5	40	8.6	1006	CNS03ESG	AL240793 Tetraodon
6	39.8	8.6	774	AL546807	AL546807 AL546807
7	39.6	8.5	457	BE943660	BE943660 UI-M-BH3-
8	39.6	8.5	494	A1848569	A1848569 UI-M-AMI-
9	39.6	8.5	1142	BM464476	BM464476 AGENCOURT
10	39.2	8.4	928	CNS00DKY	AL071865 Drosophil
11	39	8.4	850	A2673135	A2673135 ENTLI30TR
12	38.8	8.4	648	BH587629	BH587629 BOHEX88TF
13	38.6	8.3	365	A1605224	A1605224 ve78a05_x
14	38.6	8.3	388	B58764	B58764 CIT-RSP-201
15	38.6	8.3	419	BM163067	BM163067 EST565590
16	38.6	8.3	711	BF468105	BF468105 UI-M-CD0-
17	38.4	8.3	443	AQ450690	AQ450690 HS_5161_A

c	18	38.4	8.3	507	12	CNS01852	AL108896 Drosophil
	19	38.4	8.3	726	10	BM170969	BM170969 EST573492
c	20	38.4	8.3	884	12	CNS00CXG	AL060063 Drosophil
c	21	38.4	8.3	1099	12	CNS01481	AL103836 Drosophil
	22	38.4	8.3	1101	12	CNS0039G	AL063921 Drosophil
	23	38.2	8.2	641	12	AG066397	AG066397 Pan trogl
	24	38.2	8.2	907	12	CNS07COV	AL439445 T7 end of
	25	37.8	8.1	535	12	A2748844	A2748844 RPCI-24-1
c	26	37.8	8.1	878	12	CNS0187R	AL108993 Drosophil
c	27	37.8	8.1	952	12	CNS0180Y	AL108748 Drosophil
c	28	37.8	8.1	985	12	CNS00J34	AL076343 Drosophil
c	29	37.8	8.1	1023	12	CNS00CVF	AL059889 Drosophil
c	30	37.8	8.1	1101	12	CNS008X3	AL052544 Drosophil
c	31	37.6	8.1	312	12	A2821437	A2821437 2M0094009
c	32	37.6	8.1	396	12	AQ683613	AQ683613 HS_5458_B
c	33	37.6	8.1	873	12	CNS024M5	AL180950 Tetraodon
c	34	37.4	8.1	797	10	BF253656	BF253656 HVSMEF000
c	35	37.4	8.1	1007	12	CNS01413	AL104181 Drosophil
c	36	37.4	8.1	1101	12	CNS0006J	AL062049 Drosophil
c	37	37.2	8.0	353	9	AW627276	AW627276 pa28a06.y
c	38	37.2	8.0	569	10	BE886609	BE886609 601509355
c	39	37.2	8.0	607	12	AG068424	AG068424 Pan trogl
c	40	37.2	8.0	614	9	AL655660	AL655660 AL655660
c	41	37.2	8.0	761	12	BH470789	BH470789 BOGYM20TR
c	42	37.2	8.0	767	12	CNS00AQX	AL055924 Drosophil
c	43	37.2	8.0	865	12	A2675018	A2675018 ENTHS57TR
c	44	37.2	8.0	889	12	A2543817	A2543817 ENTFT53TF
c	45	37	8.0	500	9	AU087167	AU087167 AU087167

ALIGNMENTS

RESULT 1
A2539346
LOCUS ENTGP83TR Entamoeba histolytica Sheared DNA linear GSS 14-NOV-2000
DEFINITION genomic, DNA sequence.
ACCESSION A2539346
VERSION A2539346.1 GI:11145135
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica
REFERENCE 1 (bases 1 to 892)
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI:IMSS sheared DNA library
COMMENT Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI-IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 15
High quality sequence stop: 847.
Location/Qualifiers
1..892
/organism="Entamoeba histolytica".
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: PHOS1; Site:1; Bst I: Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a

[illegible]

QY	145	tccttttggatgctctctctgttaaaaggagcagctcctctggcactctgacgagaagc	204
		: : : : : : : : : : :	
Db	1049	TTTTTTKTTTCTTTTDFATTTTTTRADDFTTTTDAKDWTTTTTTAGTTTTTTTTTKDGR	990
QY	205	agctgcgcagtttacacaaaagaatggagatcagagtagctcttttgcgccaccaacgtg	264
		: : : : : : : : : : :	
Db	989	TTTTTTTTTTTTTTTATKDKKGRRRHHYMNHHYKDMGMTTHHTTTTTTMTNYNYMCV	929
QY	265	ctgagaaatttgcgtagtgcttactatcatcacacattactcttatttcacgaatatcca	324
		: : : : : : : : : : :	
Db	930	KNTKMTKMHHTTKKTKTHMTCTTTTTTYYTMMMTTKKKTKMTTNGNGNVGBTKSV	871
QY	325	cttcgcggtccgcgcgtggccgagagattg	355
		: : : : : : : : : : :	
Db	870	VSYTGMTSSSCNTKGCSTGGMGKGGKGG	840

RESULT	3
AQ421326	
LOCUS	AQ421326 681 bp DNA linear GSS 23-MAR-1999
DEFINITION	RPC1-11-167J10.TJ RPC1-11 Homo sapiens genomic clone RPC1-11-167J10, DNA sequence.
ACCESSION	AQ421326
VERSION	AQ421326.1 GI:4479050
KEYWORDS	GSS.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 681)
AUTHORS	Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter,J.C.

Unpublished (1997)
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbeet@ig.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from
Research Genet cs (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: SP6

```

1. .681
/organism="Homo sapiens"
/db_xref="GB:7563969"
/db_xref="taxon:9606"
/clone="RPCI-11-167J10"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC111 Human Male BAC Library"

```


QY	138	cattctctctttttgattgtcttcttg	165
Db	446	TATTCCTTATTTTCATTTTCTGCTG	473
RESULT	6		
LOCUS	AL546807	774 bp	linear
DEFINITION	AL546807 LTI_NFL006_PL2 Homo sapiens CDNA clone CS0D1026Y122 5 prime, mRNA sequence.	EST 16-FEB-2001	
ACCESSION	AL546807		
VERSION	AL546807.1	GI:12880281	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	Li, W.B., Gruber, C., Jessee, J., and Polayres, D.		
TITLE	Full-length cDNA libraries and normalization		
JOURNAL	Unpublished (2001)		
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.		
FEATURES	Location/Qualifiers		
source	1. .774 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="CS0D1026Y122" /clone_lib="LTI_NFL006_PL2" /tissue_type="placenta" /note="Vector: pCMVSPORT 6; Site:1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"		
BASE COUNT	188 a 56 c 111 g 318 t 101 others		
ORIGIN			
Query Match	8.6%; Score 39.8; DB 9; Length 774;		
Best Local Similarity	37.2%; Pred. No. 38;		
Matches	89; Conservative 34; Mismatches 116; Indels 0; Gaps 0;		
QY	72	ttgttagtaaaattttgctaatagtaccataatttttagtgtaggcactgtgtttatataca	131
Db	66	TTWTTTTWTATAGTGTGAWTTWKAWTTTWWGATTATTATTTTGTGTTTGTAGCAATTTAGA	125
QY	132	tataccaactctctctttttgattgtcttcttctgtttaaaggcagcactctctcttggc	191
Db	126	WAAWATTTATTTTTCWTTTWTAAWKAGATTTTATTTATAAATTTTCGTTTAAATTTTTT	185
QY	192	atctagcagaatgagctgctcagctttacacaaaagaatggagatcagactcttttg	251
Db	186	WTGCTCGTTTAAACGTTTCTGTRAGCKWTAAGAGAKTTWKATYTGVTGTTTTTTTTT	245
QY	252	tgccaccaacgtgctcgagaattgttagtggttactatcatcacactctttatttt	310
Db	246	TTTTTAAWAKTTTTTTTWTAAWAWTTTAAWAWTTTAAWAKTTTGTGTTK	304
RESULT	7		
LOCUS	BE943660	457 bp	linear
DEFINITION	UI-M-BH3-awf-g-11-0-UI.s1 NIH_BMAP_M_S4 Mus musculus cDNA clone	EST 03-OCT-2000	


```

Query Match      8.5%; Score 39.6; DB 10; Length 457;
Best Local Similarity 66.3%; Pred. No. 44;
Matches 57; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 72 ttgtgtagtaaatatttcttaagtagtaacctaatatttttagtagtagcactgtgtttataca 131
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15 TTGCAAGTCAATATTTAGCTGATGTAATTAATATTTCTCTAGGAATGTTGGTTATCCA 74

QY 132 tatatccattctctttttttgattg 157
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 75 ACTACCTATTTTCTTTTACTTTG 100

RESULT 8
LOCUS      AI848569          494 bp      mRNA      linear      EST 15-JUL-1999
DEFINITION UI-M-AM1-aga-b-02-0-UI.s2 NIH_BMAP_MAM_N Mus musculus cDNA clone
ACCESSION  AI848569
VERSION    AI848569.1 GI:5492475
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
REFERENCE  1 (bases 1 to 494)
AUTHORS   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE     Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL   Bonaldo, M.F., Lennon, G. and Soares, M.B.
MEDLINE   Normalization and subtraction: two approaches to facilitate gene
COMMENT   discovery
          Genome Res. 6 (9), 791-806 (1996)
          9704477
          Contact: Chin, H
          National Institute of Mental Health
          6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
          20852-9643, USA
          Tel: 301 443 1706
          Fax: 301 443 9890
          Email: mESP@mail.nih.gov
          The sequence contained an oligo-dT track that was present in the
          oligonucleotide that was used to prime the synthesis of first
          strand cDNA and therefore this may represent a bonafide poly A
          tail. The sequence tag present in the cDNA between the NotI site
          and the oligo-dT track served to verify it as a clone from the
          normalized amygdala library cDNA Library Preparation: M.B. Soares
          Lab Clone distribution: NIH BMAP cDNA clones will be made available
          by the means that is soon to be determined. When NIH determines the
          means for distribution of the BMAP cDNA clones, this record will be
          updated accordingly when that means is determined.
          Seq primer: M13 Forward
          POLYA=Yes.

FEATURES             Location/Qualifiers
     source           1..494
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-AM1-aga-b-02-0-UI"
                     /clone_lib="NIH_BMAP_MAM_N"
                     /dev_stage="27-32 days"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT73D-pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_MAM_N library is a normalized library constructed
                     from mouse amygdala. The tag is a string of 5 nucleotides
                     present between the Not I site and the oligo-dT track.
                     The library was constructed as described by Bonaldo,
                     Lennon and Soares, Genome Research 6: 791-806, 1996.
                     Tissue provided by Ms. Annie Novakovich, Zivic-Miller
                     Laboratories.
                     TAG_LIB=NIH_BMAP_MAM_N
                     TAG_TISSUE=amygdala
                     TAG_SEQ=GTGAG"
BASE COUNT      126 a      118 c      91 g      159 t

```

```

ORIGIN
Query Match      8.5%; Score 39.6; DB 9; Length 494;
Best Local Similarity 56.3%; Pred. No. 44;
Matches 57; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 72 ttgtgtagtaaatatttcttaagtagtaacctaatatttttagtagtagcactgtgtttataca 131
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15 TTGCAAGTCAATATTTAGCTGATGTAATTAATATTTCTCTAGGAATGTTGGTTATCCA 74

QY 132 tatatccattctctttttttgattg 157
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 75 ACTACCTATTTTCTTTTACTTTG 100

RESULT 9
LOCUS      BM464476          1142 bp      mRNA      linear      EST 05-FEB-2002
DEFINITION AGENCOURT_6438833 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5535563
ACCESSION  BM464476
VERSION    BM464476.1 GI:18513518
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
REFERENCE  1 (bases 1 to 1142)
AUTHORS   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE     Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL   NIH-MGC http://mgc.nci.nih.gov/.
COMMENT   National Institutes of Health, Mammalian Gene Collection (MGC)
          Unpublished (1999)
          Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-remail.nih.gov
          Tissue Procurement: ATCC
          cDNA Library Preparation: Life Technologies, Inc.
          DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLAM12223 row: n column: 12
          High quality sequence stop: 630.

FEATURES             Location/Qualifiers
     source           1..1142
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5535563"
                     /clone_lib="NIH_MGC_71"
                     /tissue_type="leiomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 2.1 Kb."
BASE COUNT      378 a      245 c      224 g      259 t      36 others
ORIGIN
Query Match      8.5%; Score 39.6; DB 10; Length 1142;
Best Local Similarity 48.6%; Pred. No. 41;
Matches 90; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 105 ttttagtaggtagcactgtgtttatcacatccattctcttttttagtattgtctttct 164
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 645 TTTNNNNNNNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTT 586

QY 165 gtttaatggcagctacacctctcttggcattcattagcagaatgagctgtcagtttacacaa 224
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 585 TTTTTCGACATATATACCATGTTACTTTATGTACAAATAAGGAATGGGAAGGGGAAATG 526

QY 225 aaagaatggagatcagagtactctttttgtccaccacacgtctctgagaaattttagtatt 284
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 525 AAAGAATAGAGAAACTATACGGTAGTAGTACGGTAGTGTGTGGGAACCAAAATTGCAGTTT 466

```



```

Db 124 AATTATC 130
| | | |
RESULT 12
BH587629/c
LOCUS BH587629.1 GI:17840087
DEFINITION BOHE Brassica oleracea genomic clone BOHEX88, DNA
sequence.
ACCESSION BH587629
VERSION BH587629.1
KEYWORDS GSS.
SOURCE Brassica oleracea.
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 648)
Town, C.D., Van Aken, S., Utterback, T. and Fraser, C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished (2001)
Other GSSs: BOHEX88TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
BASE COUNT 254 a 71 c 60 g 263 t
ORIGIN
1..648
/organism="Brassica oleracea"
/strain="Tol000DH3"
/db_xref="taxon:3712"
/clone="BOHEX88"
/clone_lib="BOHE"
/note="Vector: pHOS1; Site 1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHOS1 using BstXI linkers"
Query Match 8.4%; Score 38.8; DB 12; Length 648;
Best Local Similarity 50.5%; Pred. No. 65;
Matches 94; Conservative 0; Mismatches 92; Indels 0; Gaps 0;
QY 52 taacaacatctctaccacatttgtagtaaatatttgctaatagtaacctatttttag 111
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 622 TTAACAACATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 563
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 112 gtaggcactgtttatcacatatccattctcttttttttttttttttttttttttt 171
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 562 TTAGGCACAGAGATATAATATATATATATATATATATATATATATATATATATAT 503
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 172 gggcagctacctctcttgccatctagcagaatgagctgcgtcagtttacacaaaagaat 231
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 502 GGATATTTCCGAATATTTCAGGATGTTTAAAGATATCTTATCATCTATACAACT 443
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 232 ggagat 237
| | |
Db 442 AAAAAT 437
| | |
RESULT 13
AI605224/c
LOCUS AI605224.1 GI:4614381
DEFINITION ve78a05.x1 Soares_mammary_gland_NBMG Mus musculus cDNA clone
sequence.
ACCESSION AI605224
VERSION AI605224
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 388)
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 365)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:492520
This clone was previously sequenced on the 5' end only, this new
data is from the 3' end
High quality sequence stop: 357.
Location/Qualifiers
1..365
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:832304"
/clone_lib="Soares_mammary_gland_NBMG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; Site 1: Not I; Site 2: Eco
RI; 1st strand cDNA was primed with a Not I - oligo(dT)
primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCGCGAATGCTTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldi."
BASE COUNT 136 a 72 c 53 g 103 t 1 others
ORIGIN
Query Match 3.3%; Score 38.6; DB 9; Length 365;
Best Local Similarity 53.7%; Pred. No. 76;
Matches 80; Conservative 0; Mismatches 69; Indels 0; Gaps 0;
QY 47 cccataaacaatatctctaccacatttggttagtaaatatttgctaatagtaacctaat 106
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 273 CCAAAATAAATGATAAACTTTCTTTTAGGCTCTTTATTGCTTGTCTATGCTAGATT 214
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 107 tttagtaggcactgtgtttatcacatatccattctcttttttttttttttttttt 166
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 213 ATATAGTATTAACCTAAACAACTACCTGTGTCATCTTTTATTGTTGTTGTTGTTGT 154
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 167 ttaattgggcagctacctctcttgccatct 195
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 153 TTGTTGTTGTTGTTTCAAGACAGGGTTTCT 125
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
RESULT 14
B58764/c
LOCUS B58764
DEFINITION CIT-HSP-2013020.TF CIT-HSP Homo sapiens genomic clone 2013020, DNA
sequence.
ACCESSION B58764
VERSION B58764.1 GI:2613482
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 388)

```

AUTHORS Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden
,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M.
and Venter,J.C.

TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map
Building

JOURNAL Unpublished (1997)
COMMENT Other_GSSs: CIT-HSP-2013020.TR

Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html
Seq primer: M13-21
Class: BAC ends.

FEATURES

source
1. .388
/organism="Homo sapiens"
/db_xref="GDB:7042534"
/db_xref="taxon:9606"
/clone="2013020"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelOAC11; Site_1: HindIII; Site_2:
HindIII"

BASE COUNT 137 a 50 c 52 g 149 t
ORIGIN

Query Match 8.3%; Score 38.6; DB 12; Length 388;
Best Local Similarity 52.9%; Pred. No. 76;
Matches 83; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

Qy 6 aatcctaatacagatttaaaagaactctcgcgccaccattaaacaatactcc 65
Db 314 ACTAATAATCAGCTGAATTTAAATAAATCTGTGAGCAACCAATTTACATATATTA 255
Qy 66 taccaaattgtagtaaatatttgcctaagtacctaatttttagtgagcactggtt 125
Db 254 TTATAATATAATATATATAATAATAAGTTCTTATGAGCCATGTGTTCTGAATTT 195
Qy 126 tatacatatccattccctcttttttgattgcttt 162
Db 194 TTTATAGAAGAACAAATTTTCAATTTAATTCCTTTT 158

RESULT 15
BM163067/c 419 bp mRNA linear EST 04-DEC-2001
LOCUS EST565590 PYBS Plasmodium yoelii yoelii cDNA clone PYCLA4 5' end,
DEFINITION mRNA sequence.

ACCESSION BM163067.1 GI:17308748
VERSION BM163067
KEYWORDS EST.

SOURCE Plasmodium yoelii yoelii.
ORGANISM Plasmodium yoelii yoelii.
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE 1 (bases 1 to 419)
AUTHORS Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Vaidya,A.B.,
Fraser,C.M. and Carucci,D.J.

TITLE Plasmodium yoelii EST project at TIGR
JOURNAL Unpublished (2001)
COMMENT Contact: Jane Carlton

Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208

Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ATCC
http://www.malaria.mr4.org/mr4pages/index.html
Seq primer: ADF.

FEATURES

Location/Qualifiers
source
1. .419
/organism="Plasmodium yoelii yoelii"
/strain="17XL"
/db_xref="taxon:73239"
/clone="PYCLA48"
/clone_lib="PyBS"
/dev_stage="Asexual blood stages"
/lab_host="E. coli XL-1 Blue"
/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was
collected from BALB/cByJ mice infected with Py17XL
parasites, and leukocytes removed by passage over
microcrystalline cellulose columns. Total RNA was
isolated using the guanidium isothiocyanate method, and
mRNA isolated using oligo(dT)-cellulose chromatography.
First strand cDNA synthesis was completed using a 50-base
primer and reverse transcriptase in the presence of
5-methyl dCTP. After second strand synthesis, uneven
termini were treated with pfu DNA polymerase and EcoRI
adaptors ligated to the blunt ends. The sample was cleaved
with XhoI and separated on a Sephacryl S-500 column.
Size-fractionated cDNA was precipitated and ligated to
HybridZAP arms directionally using EcoRI-XhoI cleaved arms.
After packaging, the phagemid vector (PAD-GAL4) was
excised from the HybridZAP vector and plasmid DNA
isolated."

BASE COUNT 226 a 37 c 49 g 107 t
ORIGIN

Query Match 8.3%; Score 38.6; DB 10; Length 419;
Best Local Similarity 51.4%; Pred. No. 76;
Matches 89; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

Qy 85 atttgcctaagtacctaatttttagtgaggcactggtttatcacatatccattct 144
Db 277 ATATTATTATAGATAGCCATTTTCTCATTTTAATTTTGTACTTTTAGCTGTTTC 218
Qy 145 tcttttttgattgtctttctggttaaggcagctaccctcttggcacttagcagaatg 204
Db 217 ACGTTTTTTTTTTTTTTTTTTTATTTTCCCTTTACTTCTTTCTGTGATGATCTTC 158
Qy 205 agctgctgcagtttacacaaaagaatggagatcagagtactttttgtgccac 257
Db 157 ATTTATTGAAGTTTATTCATACAGAAATATATATATGATACATATGTTGTCCTC 105

Search completed: September 24, 2002, 14:45:15
Job time: 16650 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2002, 15:53:42 : Search time 3102.12 Seconds
(without alignments)
3420.162 Million cell updates/sec

Title: US-09-617-174B-1_COPY_451_957

Perfect score: 507

Sequence: 1 tggtaatacctaatacagg.....ccaagaggctgagtaggag 507

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_on.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vl.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_or.*

21: em_ov.*

22: em_pat.*

23: em_ph.*

24: em_pl.*

25: em_ro.*

26: em_sts.*

27: em_un.*

28: em_vl.*

29: em_htg_hum.*

30: em_htg_inv.*

31: em_htg_inv.*

32: em_htg_inv.*

33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	ID	Description
------------	-------------	--------	----	-------------

RESULT	1			
AC090307	AC090307	170269 bp	DNA	linear PRI 30-DEC-2001
LOCUS	Homo sapiens chromosome 18, clone RP11-851B10, complete sequence.			
DEFINITION	AC090307			
ACCESSION	AC090307			
VERSION	HTG.			
KEYWORDS	human.			
SOURCE	Homo sapiens			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	1 (bases 1 to 170269)			
AUTHORS	Birren, B., Linton, L., Nusbaum, C. and Lander, E.			
TITLE	Homo sapiens chromosome 18, clone RP11-851B10			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 170269)			
AUTHORS	Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S., Barina, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Choepel, Y., Colangelo, M., Collins, S., Collumore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,			

repeat_region	/rpt_family="AluXs" 3218. .3365
repeat_region	/rpt_family="MIR" 3565. .3590
repeat_region	/rpt_family="(TG)n" complement(3782. .4028)
repeat_region	/rpt_family="MIR" 4056. .4177
repeat_region	/rpt_family="MER5A" 4601. .5620
repeat_region	/rpt_family="HSMAR2" 5621. .5643
repeat_region	/rpt_family="(CAA)n" 5644. .5866
repeat_region	/rpt_family="HSMAR2" 6184. .6475
repeat_region	/rpt_family="L1MA2" 6677. .6837
repeat_region	/rpt_family="(TA)n" 6852. .7018
repeat_region	/rpt_family="(TA)n" 6994. .7122
repeat_region	/rpt_family="(CATATA)n" 7146. .7332
repeat_region	/rpt_family="(CATATA)n" 7371. .7526
repeat_region	/rpt_family="(TA)n" complement(7527. .7835)
repeat_region	/rpt_family="AluXs" 7936. .8080
repeat_region	/rpt_family="MTR" 8697. .8927
repeat_region	/rpt_family="AluXs" complement(9036. .9288)
repeat_region	/rpt_family="AluSq" 9873. .9934
repeat_region	/rpt_family="(GA)n" 10340. .10367
repeat_region	/rpt_family="AT_rich" 10368. .10686
repeat_region	/rpt_family="AluY" 10687. .10711
repeat_region	/rpt_family="AT_rich" complement(10870. .11043)
repeat_region	/rpt_family="L2" 11348. .11375
repeat_region	/rpt_family="(T)n" 12359. .12449
repeat_region	/rpt_family="L3" complement(12634. .12890)
repeat_region	/rpt_family="MIR" complement(13436. .13725)
repeat_region	/rpt_family="AluJb" 14244. .14297
repeat_region	/rpt_family="(CGGG)n" complement(15188. .15350)
repeat_region	/rpt_family="MIR3" 16056. .16114
repeat_region	/rpt_family="MIR3" complement(16129. .16429)
repeat_region	/rpt_family="AluXs" 16532. .16646
repeat_region	/rpt_family="MER94" 16855. .17145
repeat_region	/rpt_family="L2" 17377. .17572
repeat_region	/rpt_family="MIR3" complement(20036. .20330)
repeat_region	/rpt_family="Charlie1" complement(20333. .20486)
repeat_region	/rpt_family="Charlie1" 20625. .20691
repeat_region	/rpt_family="L2" complement(20691. .20844)

repeat_region	/rpt_family="NSRAN2"	6184. .6475
repeat_region	/rpt_family="L1MA2"	6677. .6837
repeat_region	/rpt_family="(TA)n"	6852. .7018
repeat_region	/rpt_family="(TA)n"	6994. .7122
repeat_region	/rpt_family="(CATATA)n"	7146. .7332
repeat_region	/rpt_family="(CATATA)n"	7371. .7526
repeat_region	/rpt_family="(TA)n"	7835. .7835
repeat_region	/rpt_family="(TA)n"	7936. .8080
repeat_region	/rpt_family="MIR"	8697. .8927
repeat_region	/rpt_family="AluSx"	9288. .9288
repeat_region	/rpt_family="AluSq"	9673. .9934
repeat_region	/rpt_family="(GA)n"	10340. .10367
repeat_region	/rpt_family="AT_rich"	10368. .10686
repeat_region	/rpt_family="AluY"	10687. .10711
repeat_region	/rpt_family="AT_rich"	10870. .11043
repeat_region	/rpt_family="L2"	11348. .11375
repeat_region	/rpt_family="(TY)n"	12359. .12449
repeat_region	/rpt_family="L3"	12634. .12890
repeat_region	/rpt_family="MIR"	13436. .13725
repeat_region	/rpt_family="AluJb"	14244. .14297
repeat_region	/rpt_family="(CGGGG)n"	15188. .15350
repeat_region	/rpt_family="MIR3"	16114. .16114
repeat_region	/rpt_family="MIR3"	16129. .16429
repeat_region	/rpt_family="AluSx"	16532. .16646
repeat_region	/rpt_family="MER94"	16855. .17145
repeat_region	/rpt_family="L2"	17377. .17572
repeat_region	/rpt_family="MIR3"	20036. .20330
repeat_region	/rpt_family="Charlie1"	20333. .20486
repeat_region	/rpt_family="Charlie1"	20625. .20691
repeat_region	/rpt_family="L2"	

repeat_region	6677.18837	/rpt_family="TA)n"
repeat_region	6852.7018	/rpt_family="TA)n"
repeat_region	6994.7122	/rpt_family="CATATA)n"
repeat_region	7146.7332	/rpt_family="CATATA)n"
repeat_region	7371.7526	/rpt_family="TA)n"
repeat_region	7527.7835	complement(7527.7835)
repeat_region	7936.8080	/rpt_family="AluSx"
repeat_region	8097.8927	/rpt_family="MIR"
repeat_region	9036.9288	complement(9036.9288)
repeat_region	9673.9934	/rpt_family="AluSq"
repeat_region	10340.10367	/rpt_family="(GA)n"
repeat_region	10368.10686	/rpt_family="AT_rich"
repeat_region	10687.10711	/rpt_family="AluY"
repeat_region	10870.11043	/rpt_family="AT_rich"
repeat_region	11348.11375	/rpt_family="L2"
repeat_region	12359.12449	/rpt_family="(T)n"
repeat_region	12634.12890	/rpt_family="L3"
repeat_region	13436.13725	/rpt_family="MIR"
repeat_region	14244.14297	/rpt_family="AluJb"
repeat_region	15188.15350	/rpt_family="(CGGG)n"
repeat_region	16056.16114	/rpt_family="MIR3"
repeat_region	16129.16429	complement(16129.16429)
repeat_region	16532.16546	/rpt_family="AluSx"
repeat_region	16855.17145	/rpt_family="MER94"
repeat_region	17377.17572	/rpt_family="L2"
repeat_region	20036.20330	/rpt_family="MIR3"
repeat_region	20333.20486	complement(20036.20330)
repeat_region	20625.20691	/rpt_family="Charlie"
repeat_region	20625.20691	/rpt_family="L2"

```

repeat_region      10687. 10711      /rpt_family="Alu"
repeat_region      10687. 10711      /rpt_family="Alu_rich"
repeat_region      10870. 11043      complement(10870. 11043)
repeat_region      11348. 11375      /rpt_family="L2"
repeat_region      12359. 12449      /rpt_family="(T)n"
repeat_region      12634. 12890      /rpt_family="L3"
repeat_region      13436. 13725      /rpt_family="Mir"
repeat_region      13436. 13725      /rpt_family="AluJb"
repeat_region      14244. 14297      /rpt_family="(CGGGG)n"
repeat_region      15188. 15350      complement(15188. 15350)
repeat_region      16056. 16114      /rpt_family="Mir3"
repeat_region      16129. 16429      /rpt_family="Mir3"
repeat_region      16532. 16646      /rpt_family="AluSx"
repeat_region      16855. 17145      /rpt_family="MER94"
repeat_region      17377. 17572      /rpt_family="L2"
repeat_region      20036. 20330      /rpt_family="Mir3"
repeat_region      20333. 20486      complement(20036. 20330)
repeat_region      20625. 20691      /rpt_family="Charlie"
repeat_region      20625. 20691      /rpt_family="L2"

```

repeat_region	/rpt_family="L2"	11348..11375
repeat_region	/rpt_family="(T)n"	12359..12449
repeat_region	/rpt_family="L3"	complement(12634..12890)
repeat_region	/rpt_family="MIR"	complement(13436..13725)
repeat_region	/rpt_family="AluJb"	14244..14297
repeat_region	/rpt_family="(CGGGG)n"	complement(15188..15350)
repeat_region	/rpt_family="MIR3"	16056..16114
repeat_region	/rpt_family="MIR3"	complement(16129..16429)
repeat_region	/rpt_family="AluSx"	16532..16646
repeat_region	/rpt_family="MER94"	16855..17145
repeat_region	/rpt_family="L2"	17377..17572
repeat_region	/rpt_family="MIR3"	complement(20036..20330)
repeat_region	/rpt_family="Charlie"	complement(20333..20486)
repeat_region	/rpt_family="Charlie"	20625..20691
repeat_region	/rpt_family="L2"	

```
repeat_region      /rpt_family="L3"      .12359..12449
repeat_region      complement(12634..12890)
repeat_region      /rpt_family="MIR"
repeat_region      complement(13436..13725)
repeat_region      /rpt_family="ALUJb"
repeat_region      14244..14297
repeat_region      /rpt_family="CGGGG)n"
repeat_region      complement(15188..15350)
repeat_region      /rpt_family="MIR3"
repeat_region      16056..16114
repeat_region      /rpt_family="MIR3"
repeat_region      complement(16129..16429)
repeat_region      /rpt_family="ALUSx"
repeat_region      16532..16646
repeat_region      /rpt_family="MER94"
repeat_region      16855..17145
repeat_region      /rpt_family="L2"
repeat_region      17377..17572
repeat_region      /rpt_family="MIR3"
repeat_region      complement(20036..20330)
repeat_region      /rpt_family="Charlie"
repeat_region      complement(20333..20486)
repeat_region      /rpt_family="Charlie"
repeat_region      20625..20691
repeat_region      /rpt_family="L2"
```

```
complement(12634, ".12690")
/rpt_family="MIR"
complement(13436, ".13725")
/rpt_family="AluJb"
14244, .14297
/rpt_family="(CGGGG)n"
complement(15188, ".15350")
/rpt_family="MIR3"
16056, .16114
/rpt_family="MIR3"
complement(16129, ".16429")
/rpt_family="AluSx"
16532, .16646
/rpt_family="MER94"
16855, .17145
/rpt_family="L2"
17377, .17572
/rpt_family="MIR3"
complement(20036, ".20330")
/rpt_family="Charlie1"
complement(20333, ".20486")
/rpt_family="Charlie1"
20625, .20691
/rpt_family="L2"
```

repeat_region	/rpt_family="MIR3"	complement(13436..13725)
repeat_region	/rpt_family="AluJb"	14244..14297
repeat_region	/rpt_family="(CGGGG)n"	complement(15188..15350)
repeat_region	/rpt_family="MIR3"	16056..16114
repeat_region	/rpt_family="MIR3"	complement(16129..16429)
repeat_region	/rpt_family="AluSx"	16532..16646
repeat_region	/rpt_family="MER94"	16955..17145
repeat_region	/rpt_family="L2"	17377..17572
repeat_region	/rpt_family="MIR3"	complement(20036..20330)
repeat_region	/rpt_family="Charlie1"	complement(20333..20486)
repeat_region	/rpt_family="Charlie1"	20625..20691
repeat_region	/rpt_family="L2"	

```
repeat_region      /rpt_family="MIR3"
repeat_region      /rpt_family="CGGGG)n"
repeat_region      complement(15188..15350)
repeat_region      /rpt_family="MIR3"
repeat_region      16056..16114
repeat_region      /rpt_family="MIR3"
repeat_region      complement(16129..16429)
repeat_region      /rpt_family="AluSx"
repeat_region      16532..16646
repeat_region      /rpt_family="MER94"
repeat_region      16895..17145
repeat_region      /rpt_family="L2"
repeat_region      17377..17572
repeat_region      /rpt_family="MIR3"
repeat_region      complement(20036..20330)
repeat_region      /rpt_family="Charlie1"
repeat_region      complement(20333..20486)
repeat_region      /rpt_family="Charlie1"
repeat_region      20625..20691
repeat_region      /rpt_family="L2"
```

repeat_region	/rpt_family="MIR3000"	153300
repeat_region	complement(15188..	153350)
repeat_region	/rpt_family="MIR3"	
repeat_region	16056..16114	
repeat_region	/rpt_family="MIR3"	
repeat_region	complement(16129..	16429)
repeat_region	/rpt_family="AluXs"	
repeat_region	16532..16646	
repeat_region	/rpt_family="MER94"	
repeat_region	16855..17145	
repeat_region	/rpt_family="L2"	
repeat_region	17377..17572	
repeat_region	/rpt_family="MIR3"	
repeat_region	complement(20036..	20330)
repeat_region	/rpt_family="Charlie1"	
repeat_region	complement(20333..	20486)
repeat_region	/rpt_family="Charlie1"	
repeat_region	20625..20691	
repeat_region	/rpt_family="L2"	

```
repeat_region      16056..16114  
/rpt_family="MIR3"  
repeat_region      complement(16129..16429)  
/rpt_family="AluSx"  
repeat_region      16332..16646  
/rpt_family="MER94"  
repeat_region      16855..17145  
/rpt_family="L2"  
repeat_region      17377..17572  
/rpt_family="MIR3"  
repeat_region      complement(20036..20330)  
/rpt_family="Charliel"  
repeat_region      complement(20333..20486)  
/rpt_family="Charliel"  
repeat_region      20625..20691  
/rpt_family="L2"
```

```
repeat_region      /rpt_family="L1629" .16429)
repeat_region      /rpt_family="ALUSx" .16532 .16646
repeat_region      /rpt_family="MER94" .16855 .17145
repeat_region      /rpt_family="L2" .17377 .17572
repeat_region      /rpt_family="MR3" .20330
repeat_region      /rpt_family="Charliel" .20486
repeat_region      /rpt_family="Charliel" .20625 .20691
repeat_region      /rpt_family="L2" .20625 .20691
```

```
repeat_region      16332..16646  
/rpt_family="MER94"  
repeat_region      16855..17145  
/rpt_family="L2"  
repeat_region      17377..17572  
/rpt_family="MR3"  
repeat_region      complement(20036..20330)  
/rpt_family="Charli51"  
repeat_region      complement(20333..20486)  
/rpt_family="Charlie"  
repeat_region      20625..20691  
/rpt_family="L2"
```

```
repeat_region      16855..17145  
/rpt_family="L2"  
repeat_region      17377..17572  
/rpt_family="MIR3"  
repeat_region      complement(20036..20330)  
/rpt_family="Charlie"  
repeat_region      complement(20333..20486)  
/rpt_family="Charlie"  
repeat_region      20625..20691  
/rpt_family="L2"
```

```
repeat_region 17377..17572  
/rpt_family="MIR3"  
repeat_region complement(20336..20330)  
/rpt_family="Charlie"  
repeat_region complement(20333..20486)  
/rpt_family="Charlie"  
repeat_region 20625..20691  
/rpt_family="L2"
```

```
repeat_region      complement(20036..20330)
                    /rpt_family="Charliel"
repeat_region      complement(20333..20486)
                    /rpt_family="Charliel"
repeat_region      20625..20691
                    /rpt_family="L2"
```

```
repeat_region complement(20333, .20486)
/rpt_family="Charlie"
repeat_region 20625, .20691
/rpt_family="L2"
```

```
repeat_region
20625..20691
/rpt_family="L2"
```


COMMENT

Tel:81-42-778-9923, Fax:81-42-778-9924)
On Jul 14, 2000 this sequence version replaced gi:8117315.
----- Genome Center
Center: RIKEN Genomic Sciences Center (GSC)
Center code: RIKEN
Web site: <http://hgp.gsc.riken.go.jp/>
Contact: hattori@gs.riken.go.jp
----- Project Information
Center project name: Humdraft18
Center clone name: RP11-851B10
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 171265 bases at least Q40
Consensus quality: 172100 bases at least Q30
Consensus quality: 172469 bases at least Q20
Insert size: 172836; sum-of-contigs
Quality coverage: 10.50x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 11 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

1 68896 contig of 68896 bp in length
68997 95899 contig of 26903 bp in length
96000 119268 contig of 23269 bp in length
119369 141739 contig of 23271 bp in length
141840 152655 contig of 10816 bp in length
152756 159466 contig of 6711 bp in length
159567 163465 contig of 3899 bp in length
163566 166433 contig of 2868 bp in length
166534 170112 contig of 3579 bp in length
170213 172095 contig of 1883 bp in length
172196 173836 contig of 1641 bp in length.
* NOTE: This is a 'working draft' sequence. It currently consists of 11 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
1 68896: contig of 68896 bp in length
68897 68996: gap of 100 bp
68997 95899: contig of 26903 bp in length
95900 95999: gap of 100 bp
96000 119268: contig of 23269 bp in length
119269 119368: gap of 100 bp
119369 141739: contig of 23271 bp in length
141740 141839: gap of 100 bp
141840 152655: contig of 10816 bp in length
152656 152755: gap of 100 bp
152756 159466: contig of 6711 bp in length
159467 159566: gap of 100 bp
159567 163465: contig of 3899 bp in length
163466 163565: gap of 100 bp
163566 166433: contig of 2868 bp in length
166434 166533: gap of 100 bp
166534 170112: contig of 3579 bp in length
170113 170212: gap of 100 bp
170213 172095: contig of 1883 bp in length
172096 172195: gap of 100 bp
172196 173836: contig of 1641 bp in length.

FEATURES

source

1. .173836
/organism="Homo sapiens"
/db_xref="taxon.9606"
/chromosome="18"
/map="18q22"

/map="18"
/clone="RP11-635N19"
/clone_lib="RPC1-11 Human Male BAC"
BASE COUNT 51584 a 35316 c 35660 g 50470 t 101 others
ORIGIN

Query Match 99.4%; Score 503.8; DB 2; Length 173131;
Best Local Similarity 99.6%; Pred. No. 1.5e-125;
Matches 505; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tgaagtaactcaatcacagagattttaaagaagaacttcctggccaccattataaacaata 60
Dy 52366 TGAGTAATCTTATACAGAGATTTTAAAGAAGAACTTCCTGCGCCACCCATTAAACAATA 52307
Qy 61 tctctaccattggtagtaataatttttgcataatgacctaatttttaggtaggaact 120
Dy 52306 TCTCCTACCAATTTGGTAGTAATAATTTTGTCTAATAGTACCTAATTTTAGGTAGGCACCT 52247
Qy 121 gtgtttatacatatcatcattctctctttttgattgtcttctgttttaattggtcagccta 180
Dy 52246 GTGTTTATACATATATCATTCCTCTCTTTTGTGTTGCTCTCTGTTTATGTCGAGCTA 52187
Qy 181 cctctctggcatctagcagaatgagctgctgcagtttacacaaaaagaatggagatcag 240
Dy 52186 CCTCTCTGGCATCTAGCAGAAATGAGCTGCTGCAGTTTACACAAAAAGAAATGGAGATCAG 52127
Qy 241 agtactttttgtccacaaacgtgtctgagaataattgtgttactatcatcacacatt 300
Dy 52126 AGTACTTTTGTGCGCCACCAAGCTGCTGAGAAATTTGTAGTGTACTATCATCACACATT 52067
Qy 301 actttttatcagaattatttaccctccgctcctgctggtggccgagagattgccgta 360
Dy 52066 ACTTTTATTCATCGAATATTTCACCTTCCTGCTGCTGCGGCGCGAGAGATTGCCGTA 52007
Qy 361 qcacgtctgacgtatgcatgtaactacacagccctctctgcccgaacatgtgtggaggc 420
Dy 52006 CGCATGCTGTAGTATGCATGTAACTCACAGCCCTCTCTGCGCGAACAATGTTGGAGGC 51947
Qy 421 cttttggaagctgtgcagaaacagcaacttcacgtcgaatcatctcttcaattgtgga 480
Dy 51946 CTTTGGAAAGCTGTGCACACAACTGTAACCTTCAGCCTGAAATCATCTTCTTCAATGTGGA 51887
Qy 481 caagctgccaagagctgtgagtaggag 507
Dy 51886 CAAGCTGCCAAGAGGCTTGAGTAGGAG 51860

RESULT 3
AP001404/c
LOCUS
DEFINITION Homo sapiens chromosome 18 clone RP11-851B10 map 18q22, WORKING
DRAFT SEQUENCE, 11 unordered pieces.
ACCESSION AP001404
VERSION AP001404.3 GI:9188495
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens DNA, clone:RP11-851B10.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 173,836 genomic DNA of 18q22
Published Only in Database (2000) In press
2 (bases 1 to 173836)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (14-MAR-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
(E-mail:hattori@gs.riken.go.jp, URL:<http://hgp.gsc.riken.go.jp/>).

AC015793
AC015793.2 GI:7144769
HTG: HTGS_PHASE0.
human.
Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 214212)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 18, clone RP11-1117D15
Unpublished
2 (bases 1 to 214212)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Balwin,J., Baine,N., Beckerly,R., Boguslavsky,L., Boukhgaiter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collamore,A.,
Cooke,P., DeArrellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferrelira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczy,J., Lieud,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLachlan,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tefaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 3, 2000 this sequence version replaced gi:6446894.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1590
Center clone name: 1117_D_15

* NOTE: This record contains 240 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1 695: contig of 695 bp in length
* 696 795: gap of 100 bp
* 796 1513: contig of 718 bp in length
* 1514 1613: gap of 100 bp
* 1614 2337: contig of 724 bp in length
* 2338 2437: gap of 100 bp
* 2438 3161: contig of 724 bp in length
* 3162 3261: gap of 100 bp
* 3262 3966: contig of 705 bp in length
* 3967 4066: gap of 100 bp
* 4067 4772: contig of 705 bp in length
* 4773 4872: gap of 100 bp
* 4873 5364: contig of 692 bp in length
* 5365 5664: gap of 100 bp
* 5665 6397: contig of 733 bp in length
* 6398 6497: gap of 100 bp
* 6498 7202: contig of 705 bp in length
* 7203 7302: gap of 100 bp
* 7303 8028: contig of 726 bp in length
* 8029 8128: gap of 100 bp

|||||
Db 54013 ATCACACATTTACTTTTATTTATTCAGAAATATTTCACCTTCCGGTCTGCGGGCGGAGAG 53954
Oy 351 gattccgtacgcatgtctgtacgtatcatgtactcacagcccttccgtgccgaaca 410
|||||
Db 53953 GATTGCCGTACGCATGCTGTAGTATGCATGTAACACTACAGCCCTTTCGTGCCGAACA 53894
Oy 411 tgttggagcccttttgggaagctgtgcagacacagcaacttcagctgaatcatcttt 470
|||||
Db 53993 TGTTCGAGGCGCTTTGGGAAGCTGTGCAGACAAACAGTAACCTTCAGCGTGAATTCATTCCTT 53834
Oy 471 caattgtgacaagctgccaaagagcttgagtaggag 507
|||||
Db 53833 CAATTGTGGAAGCTGCCAAGAGCGTTGAGTAGGAG 53797
RESULT 5
AX251236/c AX251236 12426 bp DNA linear PAT 05-OCT-2001
LOCUS
DEFINITION Sequence 204 from Patent WO0168912.
ACCESSION AX251236
VERSION AX251236.1 GI:15984659
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 12426)
AUTHORS Olek.A., Piepenbrock,C. and Berlin,K.
TITLE Diagnosis of diseases associated with tumor suppressor genes and
oncogenes
JOURNAL Patent: WO 0168912-A 204 20-SEP-2001;
Epigenomics AG (DE)
FEATURES
source 1. 12426
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 3791 a 137 c 2517 g 5981 t
ORIGIN
Query Match 71.4%; Score 362.2; DB 6; Length 12426;
Best Local Similarity 82.5%; Pred. No. 2.5e-87;
Matches 415; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
Oy 1 tgagtaactcctaactcacaggaattttaaaagaacacctcctgcgccaccattaaacaata 60
Db 8056 TAAATATCTCTAATCACAAAATTTTAAAAAAAACCTTCTAGCCGCCCATTTAAACAATA 7997
Oy 61 tctctaccacaaatttgtagtaataattttgctaatagtacctaatttttagtaggcact 120
Db 7996 TCTCCTACCAATTTAATAATAATATTTTACTAATAATACCTAATTTTAAATAAACACT 7937
Oy 121 gtgtttatacatatataccattcctcttttttgatgtgtcttctgtttaaaggcgagcta 180
Db 7936 ATATTATACATATATCCATCTCTTTTAAATATATCTTCTATTAATAAACAATA 7877
Oy 181 cctctctggcatctagcagaatgagctgctgcaattttacacaaaaaagaatgagatcag 240
Db 7876 CCTCTCTTAACATCTAAACAAAATAAACACTACTACAATTTACACAAAAAATAAATAACAA 7817
Oy 241 agtacttttttgcaccacacgctgtctgagaaatttgtagtattactatcatcacacatt 300
Db 7816 AATACTTTTATACCAACCAAGCTATCTAAAAAATTTAATAATTAATACTATCATCACAT 7757
Oy 301 acttttattcatogaattttccacctccggtcctgcgtggtggcgagagattgccgta 360
Db 7756 ACTTTTATTTCATCGAATATTTACCTTCCGATCCCTACGTAAACCGAAAAAATACCCTA 7697
Oy 361 cgcattgtctgcatgcatgtaactcacagcccttccctgcgcgaacatggttgagagc 420
Db 7696 CGCATATCTATACGTATACATATACTCACACCCCTTCTTACCCCGAACATATTAANAAC 7637

Oy 421 cttttggagctgtgcagacacagcaacttcagcctgaatcatctcttcaattgtgga 480
|||||
Db 7636 CTTTTAAAAACTATATAACAACAATAAATCTTCAACCTAAATCATTTCTTCAATTATAAA 7577
Oy 481 caagctgccaaagagcgttgagta 503
|||||
Db 7576 CAAACTACCAAAAAACJTAAATA 7554
RESULT 6
AX251235 AX251235 12426 bp DNA linear PAT 05-OCT-2001
LOCUS
DEFINITION Sequence 203 from Patent WO0168912.
ACCESSION AX251235
VERSION AX251235.1 GI:15984658
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 12426)
AUTHORS Olek.A., Piepenbrock,C. and Berlin,K.
TITLE Diagnosis of diseases associated with tumor suppressor genes and
oncogenes
JOURNAL Patent: WO 0168912-A 203 20-SEP-2001;
Epigenomics AG (DE)
FEATURES
source 1. 12426
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 3527 a 137 c 2591 g 6171 t
ORIGIN
Query Match 67.8%; Score 343.8; DB 6; Length 12426;
Best Local Similarity 79.9%; Pred. No. 2.3e-82;
Matches 405; Conservative 0; Mismatches 102; Indels 0; Gaps 0;
Oy 1 tgagtaactcctaactcacaggaattttaaaagaacacctcctgcgccaccattaaacaata 60
Db 4371 TGAGTAATTTTAAATTTATAGGATTTTAAAAAGAAATTTTTCGGTTATTTTAAATAATA 4430
Oy 61 tctctaccacaaatttgtagtaataattttgctaatagtacctaatttttagtaggcact 120
Db 4431 TTTTATTATTAATTTGCTAGTAAATATTTTGTAAATAGTATTTAAATTTTAGTAGGTAT 4490
Oy 121 gtgtttatacatatataccattcctcttttttgatgtgtcttctgtttaaaggcgagcta 180
Db 4491 GTGTTTATATATATATATTTTATTTTGTGATTTGTTTGTGTTTGTGTTTGTGTTTGTGTT 4550
Oy 181 cctctctggcatctagcagaatgagctgctgcaattttacacaaaaaagaatgagatcag 240
Db 4551 TTTTGTGTTTGTATTTAGTAGAATGAGTTGTTGTTGTTTATATAAANAAGAAATGAGATTAG 4610
Oy 241 agtacttttttgcaccacacgctgtctgagaaatttgtagtattactatcatcacacatt 300
Db 4611 AGTATTTTGTGTTTATTAACGTGTTTGAGAAATTTGAGTGTATTATTTATTTATATATT 4670
Oy 301 acttttattcatogaattttccacctccggtcctgcgtggtggcgagagattgccgta 360
Db 4671 ATTTTATTTTATCATCGAATATTTTATTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTT 4730
Oy 361 cgcattgtctgcatgcatgtaactcacagcccttccctgcgcgaacatggttgagagc 420
Db 4731 CGTATGTTTGTACGTATGTAATTTATATTATAGTTTTTTTTTGTTCGAATATATTGAGGT 4790
Oy 421 cttttggagctgtgcagacacagcaacttcagcctgaatcatctcttcaattgtgga 480
Db 4791 TTTTGTGGAAGTTGCTAGATAATACTAAATTTTGTAGTTTGAATTTATTTTAAATGTGGA 4850
Oy 481 caagctgccaaagagcgttgagtaggag 507
|||||
Db 4851 TAAAGTTGTTAAGAGGCTTGTAGTAGGAG 4877

RESULT 7
AC015793
LOCUS
DEFINITION
Homo sapiens chromosome 18 clone RP11-1117D15 map 18, LOW-PASS
SEQUENCE SAMPLING.
AC015793
VERSION
KEYWORDS
SOURCE
HTG: HTGS_PHASE0.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS
TITLE
JOURNAL
1 (bases 1 to 214212)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 18, clone RP11-1117D15
Unpublished
2 (bases 1 to 214212)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Castle,A., Collangelo,M., Collins,S., Collamore,A.,
Cooke,P., DeAtellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J.C., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 3, 2000 this sequence version replaced gi:6446894.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1590
Center clone name: 1117_D15

* NOTE: This record contains 240 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1
* 696 795: gap of 100 bp
* 796 1513: contig of 718 bp in length
* 1514 1613: gap of 100 bp
* 1614 2337: contig of 724 bp in length
* 2338 2437: gap of 100 bp
* 2438 3161: contig of 724 bp in length
* 3162 3261: gap of 100 bp
* 3262 3966: contig of 705 bp in length
* 3967 4066: gap of 100 bp
* 4067 4772: contig of 706 bp in length
* 4773 4872: gap of 100 bp
* 4873 5564: contig of 692 bp in length
* 5565 5664: gap of 100 bp
* 5665 6397: contig of 733 bp in length
* 6398 6497: gap of 100 bp
* 6498 7202: contig of 705 bp in length
* 7203 7302: gap of 100 bp
* 7303 8028: contig of 726 bp in length
* 8029 8128: gap of 100 bp
* 8129 8803: contig of 675 bp in length
* 8804 8903: gap of 100 bp
* 8904 9613: contig of 710 bp in length
* 9614 9713: gap of 100 bp
* 9714 10409: contig of 696 bp in length
* 10410 10509: gap of 100 bp
* 10510 11237: contig of 728 bp in length
* 11238 11337: gap of 100 bp
* 11338 12043: contig of 706 bp in length
* 12044 12143: gap of 100 bp
* 12144 12867: contig of 724 bp in length
* 12868 12967: gap of 100 bp
* 12968 13660: contig of 693 bp in length
* 13661 13760: gap of 100 bp
* 13761 14489: contig of 729 bp in length
* 14490 14583: gap of 100 bp
* 14590 15314: contig of 725 bp in length
* 15315 15414: gap of 100 bp
* 15415 16129: contig of 715 bp in length
* 16130 16229: gap of 100 bp
* 16230 16946: contig of 717 bp in length
* 16947 17046: gap of 100 bp
* 17047 17769: contig of 723 bp in length
* 17770 17869: gap of 100 bp
* 17870 18574: contig of 705 bp in length
* 18575 18674: gap of 100 bp
* 18675 19379: contig of 705 bp in length
* 19380 19479: gap of 100 bp
* 19480 20186: contig of 707 bp in length
* 20187 20286: gap of 100 bp
* 20287 20980: contig of 694 bp in length
* 20981 21080: gap of 100 bp
* 21081 21779: contig of 699 bp in length
* 21780 21879: gap of 100 bp
* 21880 22594: contig of 715 bp in length
* 22595 22694: gap of 100 bp
* 22695 23399: contig of 705 bp in length
* 23400 23499: gap of 100 bp
* 23500 24217: contig of 718 bp in length
* 24218 24317: gap of 100 bp
* 24318 25026: contig of 709 bp in length
* 25027 25126: gap of 100 bp
* 25127 25836: contig of 710 bp in length
* 25837 25936: gap of 100 bp
* 25937 26637: contig of 701 bp in length
* 26638 26737: gap of 100 bp
* 26738 27446: contig of 709 bp in length
* 27447 27546: gap of 100 bp
* 27547 28274: contig of 728 bp in length
* 28275 28374: gap of 100 bp
* 28375 29095: contig of 721 bp in length
* 29096 29195: gap of 100 bp
* 29196 29917: contig of 722 bp in length
* 29918 30017: gap of 100 bp
* 30018 30724: contig of 707 bp in length
* 30725 30824: gap of 100 bp
* 30825 31529: contig of 705 bp in length
* 31530 31629: gap of 100 bp
* 31630 32324: contig of 695 bp in length
* 32325 32424: gap of 100 bp
* 32425 33116: contig of 692 bp in length
* 33117 33216: gap of 100 bp
* 33217 33936: contig of 720 bp in length
* 33937 34036: gap of 100 bp
* 34037 34750: contig of 714 bp in length
* 34751 34850: gap of 100 bp

This clone contains STS SWSS2102 (NID:g940333).

LOCUS AC003082 170136 bp DNA linear FRI 04 FEB 2000

```
misc_feature complement(23173..23244)
/note="similar to EST AA489840 (NID:g2220715) ab04f12.s1"
23184..23261
/note="similar to EST AA366493 (NID:g2018832)"
repeat_region complement(27159..27188)
/rpt_family="L1"
repeat_region 28018..28045
/rpt_family="L1"
repeat_region complement(28082..28155)
/rpt_family="L1"
repeat_region 28086..28175
/rpt_family="L1"
repeat_region 28200..28284
/rpt_family="ALU"
repeat_region complement(29362..29403)
/rpt_family="L1"
repeat_region 29513..29694
/rpt_family="L1"
repeat_region 29943..30142
/rpt_family="L1"
repeat_region 30183..30377
/rpt_family="L1"
repeat_region 30980..31186
/rpt_family="L1"
repeat_region 31384..31422
/rpt_family="L1"
repeat_region 31441..31480
/rpt_family="L1"
repeat_region 33572..33606
/rpt_family="L1"
repeat_region 34091..34382
/rpt_family="ALU"
repeat_region complement(34850..35137)
/rpt_family="ALU"
misc_feature 35955..36007
/note="similar to EST AA425675 (NID:g2106395) zw47b05.s1"
repeat_region complement(38607..39020)
/rpt_family="L1"
repeat_region 39165..39186
/rpt_family="L1"
repeat_region complement(41821..41959)
/rpt_family="L1"
repeat_region 41957..43047
/rpt_family="L1"
repeat_region 43072..43358
/rpt_family="L1"
repeat_region 43364..43656
/rpt_family="ALU"
repeat_region 43682..44482
/rpt_family="L1"
repeat_region complement(44490..44609)
/rpt_family="L1"
repeat_region 44565..46946
/rpt_family="L1"
repeat_region complement(48709..48999)
/rpt_family="ALU"
repeat_region 49881..49901
/rpt_family="L1"
repeat_region 51668..56075
/rpt_family="L1"
misc_feature complement(52387..52447)
/note="similar to EST AA489840 (NID:g2220715) ab04f12.s1"
52387..52466
/note="similar to EST AA366493 (NID:g2018832)"
misc_feature complement(52387..52480)
/note="similar to EST AA48997 (NID:g2218599) aa54c11.s1"
56076..56419
/rpt_family="MER"
repeat_region 56420..56942
/rpt_family="L1"
repeat_region complement(57178..57226)
/rpt_family="ALU"
repeat_region 59193..59213
```

```
repeat_region /rpt_family="L1"
complement(60100..61153)
/rpt_family="MER"
repeat_region 63986..64015
/rpt_family="L1"
repeat_region 64113..64253
/rpt_family="ALU"
repeat_region complement(65648..65939)
/rpt_family="ALU"
repeat_region 66049..66338

Query Match 8.8% Score 44.4 DB 9 Length 170136
Best Local Similarity 46.7% Pred. No. 0.25
Matches 141; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

Qy 1 tgaagtaacctcaatcacaggatttttaaaagaacttctgcgcacccattaaacaata 60
Db 62665 TAAGTAATCTCACTGACATATTTCTTAGAAAAAAGATACAAATATTTTCCCTTGC 62724

Qy 61 tctctaccatattggtagtaataattttgctaaagtagtacctaaatttttagtaggcact 120
Db 62725 ACTCTATAGTGTCTCCTCAAAATATTTTAAAGTCATCTTAGTTTCTTATA 62784

Qy 121 gtgtttatcatatccattcctcttctttgttgattgttcttcttgaatgggcagcta 180
Db 62785 TTCACAATGCTTAATAAGATTTTGTAGACTCTCATGCTGAATGCATGTAATGTAGATT 62844

Qy 181 cctctcttggeatctagcagaatgagctgctgcaatttaccacaaaagaatgagatcag 240
Db 62845 CAACGTGCTAGTTTAAACACACAGCTATGCTATCTCTTTGTGAAGCTTAAATATGTG 62904

Qy 241 agtactttttgtgcacaaacgtctgtgagaattgttagtatttagtattcatcacacatt 300
Db 62905 TCTGTCTTTTATCCAATTGAGTCCCTCGTAACCTCACITGTGTAACAATATCACAGTTT 62964

Qy 301 ac 302
Db 62965 CC 62966

RESULT 10
AC084396/c
LOCUS AC084396 225053 bp DNA linear HTG 05-APR-2001
DEFINITION Homo sapiens chromosome UNK clone RP11-134N3, WORKING DRAFT
SEQUENCE, 61 unordered pieces.
ACCESSION AC084396
VERSION AC084396.3 GI:13549348
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 225053)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (30-OCT-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT
On Apr 5, 2001 this sequence version replaced gi:12863223.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0134N03
----- Summary Statistics -----
Sequencing vector: M13; 91%
Sequencing method: plasmid; 4%
```

Chemistry: Dye-primer EP; 91% of reads
Chemistry: Dye-terminator Big Dye; 4% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 184917 bases at least Q40
Consensus quality: 196549 bases at least Q30
Consensus quality: 203318 bases at least Q20
Insert size: 196000; agarose-fp
Insert size: 219053; sum-of-contigs
Quality coverage: 3.56 in Q20 bases; agarose-fp
Quality coverage: 2.91 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 61 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 4207: contig of 4207 bp in length
* 4208
* 4307: gap of unknown length
* 4308
* 10132: contig of 5825 bp in length
* 10133
* 10232: gap of unknown length
* 10233
* 15820: contig of 5588 bp in length
* 15821
* 15920: gap of unknown length
* 15921
* 20334: contig of 4414 bp in length
* 20335
* 20434: gap of unknown length
* 20435
* 27845: contig of 7411 bp in length
* 27846
* 27945: gap of unknown length
* 27946
* 32177: contig of 4232 bp in length
* 32178
* 32277: gap of unknown length
* 32278
* 38504: contig of 6227 bp in length
* 38505
* 38604: gap of unknown length
* 38605
* 47769: contig of 9165 bp in length
* 47770
* 47869: gap of unknown length
* 55641: contig of 7772 bp in length
* 55642
* 55741: gap of unknown length
* 55742
* 65119: contig of 9378 bp in length
* 65120
* 65219: gap of unknown length
* 65220
* 73704: contig of 8485 bp in length
* 73705
* 73804: gap of unknown length
* 73805
* 83677: contig of 9873 bp in length
* 83678
* 83777: gap of unknown length
* 92194: contig of 8417 bp in length
* 92195
* 92294: gap of unknown length
* 92295
* 100513: contig of 8219 bp in length
* 100514
* 106583: gap of unknown length
* 106584
* 106583: gap of unknown length
* 106584
* 117190: contig of 10507 bp in length
* 117191
* 117290: gap of unknown length
* 117291
* 118728: contig of 1438 bp in length
* 118729
* 118828: gap of unknown length
* 118829
* 120265: contig of 1437 bp in length
* 120266
* 120365: gap of unknown length
* 120366
* 120365: contig of 1287 bp in length
* 121652
* 121652: gap of unknown length
* 121753
* 122860: contig of 1108 bp in length
* 122861
* 122960: gap of unknown length
* 122961
* 124398: contig of 1338 bp in length
* 124399
* 124398: gap of unknown length
* 125617: contig of 1219 bp in length
* 125618
* 125717: gap of unknown length
* 125718
* 127326: contig of 1609 bp in length
* 127327
* 127426: gap of unknown length
* 127427
* 129643: contig of 2217 bp in length
* 129644
* 129743: gap of unknown length
* 129744
* 130956: contig of 1213 bp in length
* 130957
* 131056: gap of unknown length
* 131057
* 132284: contig of 1228 bp in length
* 132285
* 132384: gap of unknown length
* 132385
* 134025: contig of 1641 bp in length
* 134026
* 134125: gap of unknown length

134126
* 135771: contig of 1646 bp in length
* 135772
* 135772: gap of unknown length
* 135772
* 137168: contig of 1297 bp in length
* 137169
* 137268: gap of unknown length
* 137269
* 137268: contig of 1508 bp in length
* 137269
* 138776: gap of unknown length
* 138777
* 138776: gap of unknown length
* 140630: contig of 1754 bp in length
* 140631
* 140730: gap of unknown length
* 140731
* 143390: contig of 2660 bp in length
* 143391
* 143490: gap of unknown length
* 143491
* 144885: contig of 1395 bp in length
* 144886
* 144885: gap of unknown length
* 146277: contig of 1292 bp in length
* 146278
* 146377: gap of unknown length
* 146378
* 148790: contig of 2413 bp in length
* 148791
* 148890: gap of unknown length
* 148891
* 151618: contig of 2728 bp in length
* 151619
* 151718: gap of unknown length
* 151719
* 153528: contig of 1810 bp in length
* 153529
* 153628: gap of unknown length
* 153629
* 155632: contig of 2004 bp in length
* 155633
* 155732: gap of unknown length
* 155733
* 158533: contig of 2621 bp in length
* 158534
* 158533: gap of unknown length
* 161521: contig of 3068 bp in length
* 161522
* 161622: gap of unknown length
* 163929: contig of 2308 bp in length
* 163930
* 164029: gap of unknown length
* 164030
* 165640: contig of 1611 bp in length
* 165641
* 165740: gap of unknown length
* 165741
* 168699: contig of 2959 bp in length
* 168700
* 168700: gap of unknown length
* 169971: contig of 1172 bp in length
* 169972
* 170071: gap of unknown length
* 170072
* 171383: contig of 1312 bp in length
* 171384
* 171483: gap of unknown length
* 171484
* 174613: contig of 3130 bp in length
* 174614
* 174713: gap of unknown length
* 174714
* 178023: contig of 3310 bp in length
* 178024
* 178123: gap of unknown length
* 178124
* 179809: contig of 1686 bp in length
* 179810
* 179909: gap of unknown length
* 179910
* 182166: contig of 2257 bp in length
* 182167
* 182266: gap of unknown length
* 182267
* 184778: contig of 2512 bp in length
* 184779
* 184779: gap of unknown length
* 187289: contig of 2411 bp in length
* 187290
* 187389: gap of unknown length
* 187390
* 190025: contig of 2636 bp in length
* 190026
* 190125: gap of unknown length
* 190126
* 193628: contig of 3503 bp in length
* 193629
* 193729
* 193729: gap of unknown length
* 196385: contig of 2657 bp in length
* 196386
* 196485: gap of unknown length
* 196486
* 199287: contig of 2802 bp in length
* 199288
* 199387: gap of unknown length
* 199388
* 202073: contig of 2686 bp in length
* 202074
* 202074: gap of unknown length
* 202174
* 205921: contig of 3748 bp in length
* 205922
* 206021: gap of unknown length
* 206022
* 211281: contig of 5260 bp in length
* 211282
* 211381: gap of unknown length
* 211382
* 215785: contig of 4404 bp in length
* 215786
* 215885: gap of unknown length
* 215886
* 220968: contig of 5083 bp in length
* 220969
* 221088: gap of unknown length
* 225053: contig of 3985 bp in length.
* 221069

Location/Qualifiers
1. .225053
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="UNK"
/clone="RF11-134N3"

FEATURES
source

Query Match	8.4%;	Score 42.4;	DB 9;	Length 166982;	
Best Local Similarity	54.5%;	Pred. No. 0.87;	Mismatches 0;	Gaps 0;	
Matches	85;	Conservative	0;		
QY 72	tttggtagtaaatatttttctaatagtacctaatatttttagtgaggcactgctgtttatataca 131				
Db 135666	TTGTTTAGGGGATATTATGTCATATTACTTACITTTTGTGATAATCACTGTTGTAATA 135625				
QY 132	tatatcatctctctcttttttgattgtcttctgtttatggcagcagctacctctcttggc 191				
Db 135626	ACGATCCAGTAATAATTACTAGTGCTCTCCATGAAAGACAGTGTTAAAGTACTAGGG 135685				
QY 192	atctagcagaatgagctgctgcagtttacacaaaaa 227				
Db 135686	ATATAGACATAAGTCCAGCCCTTACTTCCAAAA 135721				
RESULT 13					
AB022222/c					
LOCUS	AB022222	22601 bp	DNA linear	PLN 27-DEC-2000	
DEFINITION	Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone:MUD12.				
ACCESSION	AB022222 BA000015				
VERSION	AB022222.1 GI:4159711				
KEYWORDS					
SOURCE	Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1				
ORGANISM	Arabidopsis thaliana				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.				
REFERENCE	1 (sites)				
AUTHORS	Sato,S., Nakamura,Y., Kaneko,T., Katoh,T., Asamizu,E., Kotani,H. and Tabata,S.				
TITLE	Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence features of the regions of 3,076,755 bp covered by sixty P1 and TAC clones				
JOURNAL	DNA Res. 7 (1), 31-63 (2000)				
MEDLINE	20181125				
REFERENCE	2 (bases 1 to 22601)				
AUTHORS	Nakamura,Y.				
TITLE	Direct Submission				
JOURNAL	Submitted (08-JAN-1999) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp, Tel:81-438-52-3935, Fax:81-438-52-3934)				
COMMENT	Address for correspondence: kaosekazusa.or.jp For the latest information on annotation of this clone, please see http://www.kazusa.or.jp/kaos/cgi-bin/aggd_graph.cgi?c=MUD12 Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'. The software programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, http://compbio.ornl.gov/grail-1.3/), GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html), NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and SplicePredictor (Volker Brendel, Stanford University, http://gremli1.zool.iastate.edu/cgi-bin/sp.cgi). Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, Washington University School of Medicine, St. Louis, http://genome.wustl.edu/eddy/tRNAscan-SE/). This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions. The 5' clone is MYH19 and the 3' clone is MSN9. Location/Qualifiers 1. .22601 /organism="Arabidopsis thaliana" /strain="Columbia" /db_xref="taxon:3702" /chromosome="5" /clone="MUD12" /clone_lib="Mitsui P1"				

CDS	join(1431..2080,2638..3487) /note="gene_id:MUD12.1" /codon_start=1 /evidence=not_experimental /product="protein kinase" /protein_id="BAA97351.1" /db_xref="GI:8843803" /translation="MSKQPNLESQDCSSNNNNANSPVSGSASFYSSNGSESVR LNSISLSCSSNSVSSEANLQSFDAENAFKRVFSPKPHNGNDLRWDAIQNVQ CKSNEIDLGHFRLKLGCGDIGSYLAELREMGCFAMKVMKGMGLIKRKLKRAQ TAREVILLIDHPFLPTLYSHFTEKESCLIMEPCSGDGLHILRQKPGHFSLAARF YASEVILLALEYLMHGMVVYDLKPEVMYREDGHIILSDPDLQSOFSPPTLQSTSQ PSCHIASYCIQPCIDPCKLPVACIQPCFKRFLNNRPRAKTEKASDLSPLMLIA EPTAARSMSPFVTHEYLAPEIIRGDHGSVDWMWTGFIYELLTKTKTFKNGNRET LFNVGOLKFPFEGSIFAQDLIRGLLTQDKPKRLGFKKGATEIKOHFFFNVMNAL IRSTTFEIPKPIDLSILNETLAKSSVQQQKHKSKQSDSSSGPYLDPEFF" complement(5117..5461) /note="gene_id:MUD12.2 similar to unknown protein sp1024415" /codon_start=1 /evidence=not_experimental /protein_id="BAA97352.1" /db_xref="GI:8843804" /translation="MKVVAAYLLAKLSGNENPSVADLKKIVESVGAIEIQEKIDLFSS LTKDRDVTLIAVGRKMAALSSGGGAVASGGGGAAPAEASPVSVEKKKEEKE SEDDGGNMSLFD" complement(join(7287..7481,7777..8253,8519..8755)) /note="gb AA023631.1 gene_id:MUD12.3 similar to unknown protein" /codon_start=1 /evidence=not_experimental /protein_id="BAA97353.1" /db_xref="GI:8843805" /translation="METKRVLSGSGSIDSLPLDELLSHILLSFLPTKRAASTSILSK RWRTLPFMNHLCASLDDTLIDLYPERTEEYYSFVLVGLLGRITLIFCLRVSSST LSNHAICDVNMDCGQLEELYTHIYIEDREEFNHTAPHYIAHNKLTVCYNNND QAPRLISYTPNLVYLDYSYLLNSENHFNLEARDLAFARAGRMGDHDTCL KIMNSITVQILHLSCTVEVIEQYKGRLELNQVKFLREMLEEVEKVNSTSEI ENKQLTNDILALLPKRSSKNTHVL" join(9792..10258,10626..11451,11535..11843,11943..12350, 12638..13816,14494..14763) /note="gene_id:MUD12.4" /codon_start=1 /evidence=not_experimental /product="disease resistance protein-like" /protein_id="BAA97354.1" /db_xref="GI:8843806" /translation="MATSSSCNNWYDVFPFSGEDVRKFTLSHLQLVLDRKLITSFK DNIEIRSOSIAPELVQIKDSRIAIVFSKNYSSWCLNELLEISVCRDKGOLVIP VYVALDPTHVRKOTGDMGAFERTCLNKTDEKNLARVALTHVANILGYHSAQCRANL DYVANKHLIQLTFLSTILGKONIKIDHILGALGERLKHQKVLVLLDLDQOVLNALAG QIQWFGSGSRITVVTNDKHLILSHGIENTYQVCLSKELALEMKLAFARONTPPDGF KLKAVEVVRHAGILPLGLNVILGRLNKRKYWMMLPRKGLDGIKOKARLVGDI LDNKVDIAIFRHIACLFNFEKVDIRLLADSDLNFNIGLENLVDKSLVNRNIVEM LLHQENGRIEVAQSNACEAREFLMDTICVDLDNIGTKKMLGSLDNDVIDHLEP NVHEKAFQGMRLNRLFNITPKALMSGOKIRLHPENFDYLPKLLKLLDNDKYPMRCLP SSRPERNLKLMQSELEKHWEGVSLCKLDMLEKSKNLKEIPDLSMATNLKLTIN LKYCSSLIVKISSIQLNKLTKLNMEGCTNLETLPAGINDKSLHRLDGLCCSRMRWPF DINNSIVLFDLKTSTIEFPFSLNHLKLEFDSMQMSEKLMWEGVPLTCLCKMLMSPP LAKNFNTLYLSDISPLVLPVPCGIQNLKLMELSIIRCKNLKLEPTGANFKYLDYLDLS GCKSLRFPDITSTISCLNRTGIEVPSIENFRLVYLTMLECNKILKYVSLNIFK LKLHDKADFSCGCTLTVEVSNKNTISVAATADNIQPKLLVSEASSSLCVOKSVVRFI NCPLKQCEALLQOEPVFKSLILGGEVPAYFNHRTAGNSLVIPVPTSISLDFRAL CALVDVKAMSPGRVD
-----	---

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2002, 15:48:30 : Search time 403.26 Seconds
(without alignments)
2158.596 Million cell updates/sec

Title: US-09-617-174B-1_COPY_451_957

Perfect score: 507
Sequence: 1 tgaatatactatcatcacagg.....ccaagagcgttagtagagag 507

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	507	100.0	1141	18	AAT89000	Human maspin Promo
2	362.2	71.4	12426	22	AAS46482	Tumour suppressor
3	343.8	67.8	12426	22	AAS46481	Tumour suppressor
4	38.2	7.5	1239	22	AAK68963	Human immune/haema
5	38.2	7.5	13537	22	AAK68964	Human immune/haema
6	38	7.5	56153	22	AAS46793	Tumour suppressor
7	37.4	7.4	6063	22	AAS46337	Tumour suppressor
8	37	7.3	7143	21	AAV0250	Plasmodium falcipar
9	36.8	7.3	513445	22	AAI61373	Soybean 318013 reg

10	36.6	7.2	8964	24	ABL33534	Human immune syste
11	36.6	7.2	8964	24	AAS61276	Human gene regulat
12	36.4	7.2	4421	22	AAK65974	Human immune/haema
13	36.4	7.2	61020	22	AAS46787	Tumour suppressor
14	36.2	7.1	2234	17	AAT09022	Arabidopsis thalia
15	36.2	7.1	6309	24	ABL32304	Human immune syste
16	35.8	7.1	6127	24	ABL33615	Human immune syste
17	35.8	7.1	8123	24	ABL34227	Human immune syste
18	35.6	7.0	952	19	AAK14118	H. pylori GHP0.122
19	35.6	7.0	9293	22	AAS45320	Chemically pretrea
20	35.4	7.0	7403	22	AAS46804	Tumour suppressor
21	35.4	7.0	7403	24	ABL34217	Human immune syste
22	35.4	7.0	56153	22	AAS46794	Tumour suppressor
23	35.2	6.9	11047	22	AAS45480	Chemically pretrea
24	35.2	6.9	11047	24	ABL33985	Human immune syste
25	35.2	6.9	28091	22	ABAI5375	Human nervous syste
26	35.2	6.9	28091	22	AAK77432	Human immune/haema
27	35.2	6.9	28120	22	ABAI5379	Human nervous syste
28	35.2	6.9	28120	22	AAK77434	Human immune/haema
29	35.2	6.9	114793	22	AAK08215	Human genome from
30	35	6.9	15592	22	AAS46454	Tumour suppressor
31	35	6.9	15592	24	ABL33327	Human immune syste
32	35	6.9	16509	24	ABL33321	Human immune syste
33	34.8	6.9	1494	20	AAK20170	Enterococcus faeca
34	34.8	6.9	6134	24	ABL33183	Human immune syste
35	34.8	6.9	6259	20	AAK13066	Enterococcus faeca
36	34.8	6.9	7069	22	AAS46654	Tumour suppressor
37	34.8	6.9	7069	24	ABL33353	Human immune syste
38	34.8	6.9	7069	24	AAS61220	Human gene regulat
39	34.6	6.8	370	22	AAI64576	Human polynucleoti
40	34.6	6.8	890	22	AAI15375	Human breast cance
41	34.6	6.8	1192	22	ABAI9853	Human nervous syste
42	34.6	6.8	6136	24	AAS61297	Human gene regulat
43	34.6	6.8	18028	23	ABL33284	Drosophila melanog
44	34.6	6.8	25955	22	ABAI9852	Human nervous syste
45	34.4	6.8	834	22	AAH03576	Human cDNA clone (

ALIGNMENTS

RESULT 1	
AAT89000	standard; cDNA: 1141 BP.
AC	AAI89000:
XX	22-APR-1998 (first entry)
XX	Human maspin promoter and partial coding cDNA.
DE	
XX	Masspin: serpin; mammary epithelial cell; human; promoter; malignant;
KW	tumour cell; treatment; prostate cancer; breast cancer; metastasis; ss.
XX	
OS	Homo sapiens.
XX	
FT	key
FT	promoter
FT	Location/Qualifiers
FT	1..1141
FT	/*tag- a
FT	/note- "maspin promoter region"
FT	443..449
FT	/*tag- b
FT	/note- "Putative regulatory element AP2"
FT	451..457
FT	/*tag- c
FT	/note- "Putative regulatory element AP1"
FT	485..490
FT	/*tag- d
FT	/note- "Putative regulatory element Ets"
FT	684..691
FT	/*tag- e
FT	/note- "Putative regulatory element HRE"
FT	846..851
FT	misc_signal

```

FT      /tag= f
FT      /note= "Putative regulatory element Ets"
FT      misc_signal
FT      907..913
FT      /tag= g
FT      /note= "Putative regulatory element AP1"
FT      prim_transcript 958..1141
FT      /tag= h
FT      /note= "partial"
XX
XX      MO9736179-A1.
XX
XX      02-OCT-1997.
XX
XX      28-MAR-1997; 97WO-US05186.
XX
XX      28-MAR-1996; 96US-0014368.
XX
XX      (DAND ) DANA FARBER CANCER INST INC.
XX      (DAND ) DANA FARBER CANCER INST.
XX      (PARD/) PARDEE A.
XX
XX      Sagar R, Zhang M;
XX      WPI; 1997-489785/45.
XX
XX      Maspin gene promoter fragment - used to identify compounds for
XX      treatment of prostate or breast cancer
XX
XX      Claim 4; Fig 3; 51pp; English.
XX
XX      This sequence encodes the human maspin promoter region, including the
XX      transcription start site. Maspin is a serpin which is expressed in
XX      mammary epithelial cells. Its expression in these cells decreases
XX      with increasing malignancy and is lost in during metastasis. Maspin
XX      protein is also known to inhibit the mobility of tumour cells. This gene
XX      can be used in method for screening compounds to identify candidate
XX      compounds for the treatment of prostate cancer, or breast cancer. It can
XX      also be used to identify compounds that increase the expression of
XX      maspin, and for detecting the presence of metastatic prostate epithelial
XX      cells.
XX
XX      Sequence 1141 BP; 261 A; 292 C; 277 G; 311 T; 0 other;
XX

```

```

Query Match      100.0%; Score 507; DB 18; Length 1141;
Best Local Similarity 100.0%; Pred. NO.2.9e-135;
Matches 507; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 tgagtaaccctaatacagagatttaaaagaactctctgagccaccattaaacata 60
DB      451 tgagtaaccctaatacagagatttaaaagaactctctgagccaccattaaacata 510
OY      61 tctctaccgaatttggtgtaataatttggtaataagacctaattttagtgagcact 120
DB      511 tctctaccgaatttggtgtaataatttggtaataagacctaattttagtgagcact 570
OY      121 ggtttatacatatatacattcctctttttgattgctgcttcgtttaatgagcgagcta 180
DB      571 ggtttatacatatatacattcctctctttttgattgctgcttcgtttaatgagcgagcta 630
OY      181 cctctcttgcatctagcagaatgagctgctgagtttaacacaaaagaatgagagatag 240
DB      631 cctctcttgcatctagcagaatgagctgctgagtttaacacaaaagaatgagagatag 690
OY      241 agtaacttttggcaccacaacgctgctgagaattttagtgatctatcatcacacatt 300
DB      691 agtaacttttggcaccacaacgctgctgagaattttagtgatctatcatcacacatt 750
OY      301 acttttaattcatcgaaatttcacacttcctgctgctgagcgagagagattgcgla 360
DB      751 acttttaattcatcgaaatttcacacttcctgctgctgagcgagagagattgcgla 810
OY      361 cgcattctgtatgcatgcatgtaactacagacccttcctgctgcccgaacattgttgagcc 420

```

```

DB      811 cgcattctgtatgcatgcatgtaactacagacccttcctgctgcccgaacattgttgagcc 870
OY      421 ctttgaagctgtgcaacaacagcaacttcacactggaatcactcttccaattgcyga 480
DB      871 ctttgaagctgtgcaacaacagcaacttcacactggaatcactcttccaattgcyga 930
OY      481 caagctgcccaagagagcttgatagag 507
DB      931 caagctgcccaagagagcttgatagag 957

RESULT 2
AAS46482/C
ID      AAS46482 standard; DNA; 12426 BP.
XX
XX      AAS46482;
XX
XX      18-DEC-2001 (first entry)
XX
XX      Tumour suppressor gene derived chemically modified sequence #204.
XX
XX      Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
XX      cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
XX      cytosine methylation; ds.
XX
XX      Homo sapiens.
XX
XX      WO200168912-A2.
XX
XX      20-SEP-2001.
XX
XX      15-MAR-2001; 2001WO-EP02955.
XX
XX      15-MAR-2000; 2000DE-1013847.
XX      06-APR-2000; 2000DE-1019058.
XX      07-APR-2000; 2000DE-1019173.
XX      30-JUN-2000; 2000DE-1032529.
XX      01-SEP-2000; 2000DE-1043826.
XX
XX      (EPIC-) EPIGENOMICS AG.
XX
XX      Olek A, Piepenbrock C, Berlin K;
XX
XX      WPI; 2001-602752/68.
XX
XX      Fragments of chemically modified genes associated with tumour suppressor
XX      genes and oncogenes, useful in designing primers and probes for
XX      analysing diseases associated with cytosine methylation state e.g.
XX      cancer
XX
XX      Claim 1; SEQ ID No 204; 27pp; English.
XX
XX      The invention relates to a nucleic acid comprising a sequence of 18
XX      bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
XX      bisulphite, of genes associated with tumour suppression and
XX      oncogenes having a sequence taken from 536 (actually 533 since
XX      numbers 408, 458 and 500 are missing from the sequence listing) sequences
XX      (5S) and sequences complementary to (5S). The nucleic acid may be a
XX      peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
XX      form part of a set of probes for detecting the cytosine methylation state
XX      and/or single nucleotide polymorphisms and also to be used in an
XX      array for analysing diseases associated with CpG dinucleotides e.g.
XX      cancers and tumours. The probes can also be used in a method for
XX      ascertaining genetic and/or epigenetic parameters for the diagnosis
XX      and/or therapy of existing diseases or the predisposition to specific
XX      diseases, by analysing cytosine methylations. The parameters may be
XX      compared to another set of genetic and/or epigenetic parameters, the
XX      differences serving as basis for diagnosis and/or prognosis events which
XX      are disadvantageous to patients. The present sequence is one of the
XX      533 genomic sequences derived from tumour suppressor genes and
XX      oncogenes. Sequences with even numbered Seq ID numbers are the
XX      complementary sequence of the corresponding odd numbered sequence (e.g.

```


CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence
CC is missing).
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pat_sequences.
XX
XX Sequence 12426 BP: 3791 A; 137 C; 2517 G; 5961 T; 0 other;

Query Match	71.48;	Score 362.2;	DB 22;	Length 12426;
Best Local Similarity	82.58;	Pred. No. 2.1e-93;		
Matches 415; Conservative	0;	Mismatches 88;	Indels 0;	Gaps 0;

[illegible]

RESULT	3
AAS46481	
ID	AAS46481 standard; DNA; 12426 BP.
XX	
AC	
XX	AAS46481;
DT	18-DEC-2001 (first entry)
DE	Tumour suppressor gene derived chemically modified sequence #203.
XX	
KM	Human: tumour suppressor gene; oncogene; antitumour; cytostatic;
KW	cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KX	cytosine methylation; ds.
XX	
OS	Homo sapiens.
XX	
PN	MO200168912-A2.
XX	
PD	20-SEP-2001.
XX	
PF	15-MAR-2001; 2001WO-EP02955.
XX	

PR 15-MAR-2000; 2000DE-1013847.
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
XX
PA (EPiG-) EPIGENOMICS AG.
XX
PI Olek A., Pletzenbrock C., Berlin K;
XX
DR WPI; 2001-602752/68.

PT Fragments of chemically modified genes associated with tumour suppressor
PT genes and oncogenes, useful in designing primers and probes for
PT analysing diseases associated with cytosine methylation state e.g.
PT cancer -

PS Claim 1; SEQ ID No 203; 27pp; English.

The invention relates to a nucleic acid comprising a sequence of 18 bases, of a segment of chemically preterreated DNA (Cp DNA) e.g. with bisulphite, of genes associated with tumour suppression and oncogenes having a sequence taken from 536 (actually 533 since numbers 408, 458 and 500 are missing from the sequence listing) sequences (5S) and sequences complementary to (5S). The nucleic acid may be a peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may form part of a set of probes for detecting the cytosine methylation states and/or single nucleotide polymorphisms and also to be used in an array for analysing diseases associated with CpG dinucleotides e.g. cancers and tumours. The probes can also be used in a method for ascertaining genetic and/or epigenetic parameters for the diagnosis and/or therapy of existing diseases or the predisposition to specific diseases, by analysing cytosine methylations. The parameters may be compared to another set of genetic and/or epigenetic parameters, the differences serving as basis for diagnosis and/or prognosis events which are disadvantageous to patients. The present sequence is one of the 533 genomic sequences derived from tumour suppressor genes and oncogenes.

CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC http://www.wipo.int/pub/published_pct/sequences.

Sequence 12426 BP; 3527 A; 137 C; 2591 G; 6171 T; 0 other;

Query Match	67.88	Score 343.8	DB 22	Length 12426
Best Local Similarity	79.98	Pred. NO. 4e-88		
Matches 405		Conservative	0	Mismatches 102
				Indels 0
				Gaps 0

QY	1	tgtgtaactactaacagagatttttaaaaagaactcctctgagccaccatttaacata	60
Db	4371	tggaaattcttaataatagaattcttaaaaaagaattcttcgtattattataataata	4430
QY	61	tctctctacccaattctgttagtaaaatattctgtcaataagtaaccaaattcttagtaagcatt	120
Db	4431	ttttttattattctgtgagaaattcttctgaataagtaattcaattcttagtgattat	4490
QY	121	gtgttttaacatataaccattcctctcttttttgattgctcttcctctttaaaggagata	180
Db	4491	gtgtttacataataataattatctctctctctctctgtattctcttcgtcaatgggaaagta	4550
QY	181	cctctctctgtgacattcagcagaaatgaagctgcgcagattacacaaaagaatctgaagatcag	240
Db	4551	ttctctctctgtactctgtagaaatgaagctctctctgtactctacaanaagaatcgagattag	4610
QY	241	agtaactttttgtgcaccaacgcgtctctgtgaaattgtgtatctatacaacaact	300
Db	4611	agtaactttctgtgtattaacgctgcttggaaattctgtagtctatattatataatct	4670
QY	301	acttaattcacaagaatattacactctccgcgtcctgcgtgagccgaagagattctgcga	360
Db	4671	attcttaattcacaagaatattatctcttcgcgtctctgcgtgagccgaagagattctgcga	4730

RESULT	4
AAK68963	
ID	AAK68963 standard; DNA: 1239 BP.
XX	
AC	AAK68963;
XX	
D7	06-NOV-2001 (first entry)
XX	
DE	Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23775.
XX	
KW	Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW	cytostatic; gene therapy; vaccine; metastasis; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200157182-A2.

RESULT	4
AAK68963	
ID	AAK68963 standard; DNA; 1239 BP.
XX	
AC	AAK68963;
XX	
DT	06-NOV-2001 (first entry)
XX	
DE	Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23775.
XX	
KW	Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KX	cytostatic; gene therapy; vaccine; metastasis; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200157182-A2.
XX	
PD	09-AUG-2001.
XX	
PF	17-JAN-2001; 2001WO-US01354.
XX	
PR	31-JAN-2000; 2000US-0179065.
PR	04-FEB-2000; 2000US-0180628.
PR	24-FEB-2000; 2000US-0184664.
PR	02-MAR-2000; 2000US-0186350.
PR	16-MAR-2000; 2000US-0189874.
PR	17-MAR-2000; 2000US-0190076.
PR	18-APR-2000; 2000US-0198123.
PR	19-MAY-2000; 2000US-0205515.
PR	07-JUN-2000; 2000US-0209467.
PR	28-JUN-2000; 2000US-0214886.
PR	30-JUN-2000; 2000US-0215135.
PR	07-JUL-2000; 2000US-0216647.
PR	07-JUL-2000; 2000US-0216880.
PR	11-JUL-2000; 2000US-0217487.
PR	11-JUL-2000; 2000US-0217496.
PR	14-JUL-2000; 2000US-0218290.
PR	26-JUL-2000; 2000US-0220963.
PR	26-JUL-2000; 2000US-0220964.
PR	14-AUG-2000; 2000US-0224516.
PR	14-AUG-2000; 2000US-0224519.
PR	14-AUG-2000; 2000US-0224519.
PR	14-AUG-2000; 2000US-0225213.
PR	14-AUG-2000; 2000US-0225214.
PR	14-AUG-2000; 2000US-0225266.
PR	14-AUG-2000; 2000US-0225267.
PR	14-AUG-2000; 2000US-0225267.
PR	14-AUG-2000; 2000US-0225268.
PR	14-AUG-2000; 2000US-0225270.
PR	14-AUG-2000; 2000US-0225447.
PR	14-AUG-2000; 2000US-0225757.
PR	14-AUG-2000; 2000US-0225758.
PR	14-AUG-2000; 2000US-0225759.
PR	18-AUG-2000; 2000US-0226279.
PR	22-AUG-2000; 2000US-0226681.
PR	22-AUG-2000; 2000US-0226688.
PR	22-AUG-2000; 2000US-0227182.
PR	23-AUG-2000; 2000US-0227009.
PR	30-AUG-2000; 2000US-0228524.
PR	01-SEP-2000; 2000US-0229287.
PR	01-SEP-2000; 2000US-0229343.

PR	17-SEP-2000	2000US-0229344
PR	01-SEP-2000	2000US-0229345
PR	05-SEP-2000	2000US-0229509
PR	05-SEP-2000	2000US-0229513
PR	05-SEP-2000	2000US-0230437
PR	06-SEP-2000	2000US-0230438
PR	06-SEP-2000	2000US-0231242
PR	08-SEP-2000	2000US-0231243
PR	08-SEP-2000	2000US-0231244
PR	08-SEP-2000	2000US-0231413
PR	08-SEP-2000	2000US-0231414
PR	08-SEP-2000	2000US-0232081
PR	08-SEP-2000	2000US-0232081
PR	12-SEP-2000	2000US-0231968
PR	14-SEP-2000	2000US-0232397
PR	14-SEP-2000	2000US-0232398
PR	14-SEP-2000	2000US-0232399
PR	14-SEP-2000	2000US-0232401
PR	14-SEP-2000	2000US-0232401
PR	14-SEP-2000	2000US-0233063
PR	14-SEP-2000	2000US-0233064
PR	14-SEP-2000	2000US-0233065
PR	14-SEP-2000	2000US-0234223
PR	21-SEP-2000	2000US-0234423
PR	21-SEP-2000	2000US-0234423
PR	25-SEP-2000	2000US-0234997
PR	25-SEP-2000	2000US-0234998
PR	25-SEP-2000	2000US-0235484
PR	26-SEP-2000	2000US-0235834
PR	27-SEP-2000	2000US-0235836
PR	27-SEP-2000	2000US-0235837
PR	29-SEP-2000	2000US-0236367
PR	29-SEP-2000	2000US-0236368
PR	29-SEP-2000	2000US-0236369
PR	29-SEP-2000	2000US-0236370
PR	29-SEP-2000	2000US-0236370
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237038
PR	02-OCT-2000	2000US-0237039
PR	02-OCT-2000	2000US-0237039
PR	02-OCT-2000	2000US-0237040
PR	13-OCT-2000	2000US-0239343
PR	13-OCT-2000	2000US-0239343
PR	20-OCT-2000	2000US-0240960
PR	20-OCT-2000	2000US-0241221
PR	20-OCT-2000	2000US-0241785
PR	20-OCT-2000	2000US-0241786
PR	20-OCT-2000	2000US-0241787
PR	20-OCT-2000	2000US-0241808
PR	20-OCT-2000	2000US-0241809
PR	20-OCT-2000	2000US-0241826
PR	01-NOV-2000	2000US-0244617
PR	08-NOV-2000	2000US-0246474
PR	08-NOV-2000	2000US-0246475
PR	08-NOV-2000	2000US-0246476
PR	08-NOV-2000	2000US-0246477
PR	08-NOV-2000	2000US-0246478
PR	08-NOV-2000	2000US-0246523
PR	08-NOV-2000	2000US-0246524
PR	08-NOV-2000	2000US-0246525
PR	08-NOV-2000	2000US-0246526
PR	08-NOV-2000	2000US-0246527
PR	08-NOV-2000	2000US-0246528
PR	08-NOV-2000	2000US-0246532
PR	08-NOV-2000	2000US-0246539
PR	08-NOV-2000	2000US-0246610
PR	08-NOV-2000	2000US-0246611
PR	08-NOV-2000	2000US-0246613
PR	17-NOV-2000	2000US-0249207
PR	17-NOV-2000	2000US-0249208
PR	17-NOV-2000	2000US-0249210
PR	17-NOV-2000	2000US-0249211
PR	17-NOV-2000	2000US-0249211
PR	17-NOV-2000	2000US-0249213

[illegible]

PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.

PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232337.
 PR 14-SEP-2000; 2000US-0232338.
 PR 14-SEP-2000; 2000US-0232339.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.

PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 11-DEC-2000; 2000US-0251990.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM:
 XX
 XX WPI: 2001-483426/52.
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and
 XX metastasis -
 PS Disclosure; SEQ ID NO 23776; 3071pp + Sequence Listing; English.
 XX
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patient's own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/hematopoietic-related diseases, especially
 CC cancers and cancer metastases of haematopoietic antigen genomic
 CC sequences from the present invention. AAK4942 to AAK54950 and AAM82169
 CC represent sequences used in the exemplification of the present invention.
 XX
 XX Sequence 13537 BP; 4245 A; 2216 C; 2328 G; 4748 T; 0 other;

Query Match 7.5%; Score 38.2; DB 22; Length 13537;
 Best Local Similarity 48.4%; Pred. No. 1.4;
 Matches 106; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

QY 1 tgagtaacctaatcacggatttaaaagaacttcgcggcaccattaaacata 60
 |||||
 DB 3161 ttgatttattcataagaacattgttaagtaattcttattcaaaagcttaacat 3220
 |||||
 QY 61 tctctaccaattgtagtaataatttgctaatagtacttaatttagtgaccat 120
 |||||
 DB 3221 ttccaattaaattgacaacacttcttacaagaataatgataatcgtatagct 3280
 |||||
 QY 121 gtgttatacatatataccattctctttttgattgctcttcgttaatggcagcta 180
 |||||
 DB 3281 ttcatatgatacaaatgltagccttattttgtttgttttaagcattggcagctg 3340
 |||||
 QY 181 cctctcttggaatcagcgaatgagcgtgcgcaattta 219
 |||||
 DB 3341 ttgtgaagtcataagaagcgagtaaggaattta 3379
 |||||

01	RESULT	6	
02	AA546793		
03	ID	AA546793	standard; DNA: 56153 BP.
04	XX		
05	AC	AA546793;	
06	XX		
07	DT	18-DEC-2001	(first entry)
08	DE		Tumour suppressor gene derived chemically modified sequence #519.
09	XX		
10	KW	Human: tumour suppressor gene; oncogene; antitumour; cytostatic;	
11	KM	cancer; tumour; Cpg dinucleotide; single-nucleotide polymorphism; SNP;	
12	XX	cytosine methylation; ds.	
13	OS	Homo sapiens.	
14	XX		
15	PN	WO200168912-A2.	
16	PD		
17	XX	20-SEP-2001.	
18	XX		
19	PF	15-MAR-2001; 2001WO-EP02955.	
20	XX		
21	PR	15-MAR-2000; 2000DE-1013847.	
22	PR	06-APR-2000; 2000DE-1019058.	
23	PR	07-APR-2000; 2000DE-1019173.	
24	PR	30-JUN-2000; 2000DE-1032529.	
25	PR	01-SEP-2000; 2000DE-1043826.	
26	XX		
27	PA	(EPIC-) EPIGENOMICS AG.	
28	XX		
29	P1	Olek A, Piepenbrock C, Berlin K;	
30	DR	WPI: 2001-602752/68.	
31	XX		
32	PT	Fragments of chemically modified genes associated with tumour suppressor	
33	PT	genes and oncogenes, useful in designing primers and probes for	
34	PT	analysing diseases associated with cytosine methylation state e.g.	
35	PT	cancer	
36	XX		
37	PS	Claim 1; SEQ ID No 519; 27pp; English.	
38	CC		
39	CC	The invention relates to a nucleic acid comprising a sequence of 18	
40	CC	bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with	
41	CC	bisulphite, of genes associated with tumour suppression and	
42	CC	oncogenes having a sequence taken from 536 (actually 533 since	
43	CC	numbers 408, 458 and 500 are missing from the sequence listing) sequences	
44	CC	(SS) and sequences complementary to (Ss). The nucleic acid may be a	
45	CC	peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may	
46	CC	form part of a set of probes for detecting the cytosine methylation state	
47	CC	and/or single nucleotide polymorphisms and also to be used in an	
48	CC	array for analysing diseases associated with Cpg dinucleotides e.g.	
49	CC	cancers and tumours. The probes can also be used in a method for	
50	CC	ascertaining genetic and/or epigenetic parameters for the diagnosis	
51	CC	and/or therapy of existing diseases or the predisposition to specific	
52	CC	diseases, by analysing cytosine methylations. The parameters may be	
53	CC	compared to another set of genetic and/or epigenetic parameters, the	
54	CC	differences serving as basis for diagnosis and/or prognosis events which	
55	CC	are disadvantageous to patients. The present sequence is one of the	
56	CC	533 genomic sequences derived from tumour suppressor genes and	
57	CC	oncogenes.	
58	CC	Note: The sequence data for this patent did not form part	
59	CC	of the printed specification, but was obtained in electronic	
60	CC	format directly from WIPO at	
61	CC	ftp.wipo.int/pub/published_pcl_sequences.	
62	XX		
63	SO	Sequence 56153 BP: 13943 A; 1002 C; 14095 G; 27113 T; 0 other;	
64			
65	Query Match	7.5%; Score 38; DB 22; Length 56153;	
66	Best Local Similarity	47.5%; Pred. No. 2.8; 125; Indels	
67	Matches	113; Conservative	
68			
69	73	tttgtagtaataatttctgaatagaccataattttagtagtaggcactggttatacat	132

Db 31726 ttgggataatagatgatgtacattcttcttagtttttcgttttttgatcttatctt 31785
Oy 133 ataccatccctcctcttttttgatgtctcttcgcgttaaggcgagctacccctctggca 192
Db 31786 ttaatgagtcttttttgatatagtttagtttaattcttcgttaatttttttttttgtt 31845
Oy 193 tcctagaagaatgagctgtcgttcgaattacaacaaaagaatggagalcagagacttttgc 252
Db 31846 ttctgtagtaataatttatttatttattgttttagaagaactttaagaggtctaataattt 31905
Oy 253 gccaccaacggtctcgtagaatctgttagttgtacctacatcacaaacttattt 310
Db 31906 ttgaagttatttttttttagtttttttttagtttattgtaaatgltacgtagttttttt 31963

RESULT 7
AAS46337
ID AAS46337 standard; DNA; 6063 BP.
XX
AC AAS46337;
XX
DT 18-DEC-2001 (first entry)
DE Tumour suppressor gene derived chemically modified sequence #59.
XX
KM Human: tumour suppressor gene; oncogene; antitumour; cytostatic;
KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
XX cytosine methylation; ds.
OS Homo sapiens.
XX
PN WO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP02955.
XX
PR 15-MAR-2000; 2000DE-1013847.
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602752/68.
XX
PT Fragments of chemically modified genes associated with tumour suppressor
PT genes and oncogenes, useful in designing primers and probes for
PT analysing diseases associated with cytosine methylation state e.g.
PT cancer -
PS
XX Claim 1; SEQ ID No 59; 27pb; English.

The invention relates to a nucleic acid comprising a sequence of 18 bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with bisulphite, of genes associated with tumour suppression and oncogenes having a sequence taken from 536 (actually 533 since numbers 408, 458 and 500 are missing from the sequence listing) sequences (Ss) and sequences complementary to (Ss). The nucleic acid may be a peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may form part of a set of probes for detecting the cytosine methylation state and/or single nucleotide polymorphisms and also to be used in an array for analysing diseases associated with CpG dinucleotides e.g. cancers and tumours. The probes can also be used in a method for ascertaining genetic and/or epigenetic parameters for the diagnosis and/or therapy of existing diseases or the predisposition to specific diseases, by analysing cytosine methylations. The parameters may be compared to another set of genetic and/or epigenetic parameters, the differences serving as basis for diagnosis and/or prognosis events which

XX New purified nucleic acid for producing a soybean plant having soybean
 PT cyst nematode resistance and for use in plant breeding programs -
 XX
 PS Claim 30: Page 596-893; 1353pp; English.
 XX
 CC The invention relates to nucleic acid molecules from regions of the
 CC soybean genome which are associated with soybean cyst nematode (SCN)
 CC resistance. The nucleic acids are used to transform plants, and can
 CC produce soybean plants having an rhg1 or an Rhg4 SCN resistant allele.
 CC The nucleic acids can be used for investigating rhg1 or Rhg4 haplotypes
 CC of soybean plants and for introgressing SCN resistance or partial SCN
 CC resistance into soybean plants. They can also be used in plant breeding
 CC programmes. The invention also relates to proteins encoded by such
 CC nucleic acid molecules, as well as antibodies capable of recognising
 CC these proteins. The present sequence is a nucleic acid molecule
 CC provided in the specification.
 CC
 SQ Sequence 513445 BP; 173367 A; 85402 C; 83912 G; 170492 T; 272 other;

Query Match
 Best Local Similarity 7.3%; Score 36.8; DB 22; Length 513445;
 Matches 130; Conservative 0; Mismatches 137; Indels 1; Gaps 1;

OY 56 caatatccctacccaattggtgagtaaatattgtgctaagtaactatttttaggtag 115
 DB 336067 CCATAGTCATCTACTTCTTACAAATTCCTCTCAATTTGGAAATACCTTTTATTTCT 336008
 OY 116 gcaatggtttacacatacatccatcctcttctttagatgcttcttcttgaatgagc 175
 DB 336007 GCTTAGAGTTTGGAAAAATATATCTACTGTTATGTTGTTGATGGAATGAAGG 335948
 OY 176 agctactctcttgagatcagag-aatgagctgctgcaagttacacaaaagaatga 234
 DB 335947 ACTCATATCCCTGTGTAAGAAATTTAGTTGATCAGTTTCAAGTATCAAAAAGAAATCT 335888
 OY 235 gatcagagactcttctgcccacacagctcgtcgaagaattttagtctacatcata 294
 DB 335887 CACGAGAGATTTACTGTTTATATATGTGCGCAATGATTAAGTTCACTTTATCG 335828
 OY 295 cacattacttattcatcagaatatt 322
 DB 335827 CCCAGAAAAATATTCGAAAAAATATTT 335800

RESULT 10
 ABL33534
 ID ABL33534 standard; DNA; 8964 BP.
 AC ABL33534;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Human immune system associated gene SEQ ID NO: 1507.
 XX
 KW Human: immune system disease; cytosine methylation; antiasthmatic;
 KW antileukemic; antianemic; cytosinetic; noctropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antineuritic; antileukemic; antidiabetic; antipsoriatic;
 KW antileukemic; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene: ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200200928-A2.
 XX
 PD 03-JAN-2002.
 XX
 PF 02-JUL-2001; 2001WO-EP07537.
 XX

PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Plepenbrock C, Berlin K;
 DR WPI: 2002-130909/17.
 XX
 PT Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation -
 PS Claim 1: SEQ ID NO 1507; 32pp + Sequence Listing; German.
 XX
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 CC
 SQ Sequence 8964 BP; 2596 A; 89 C; 1815 G; 4464 T; 0 other;

Query Match
 Best Local Similarity 7.2%; Score 36.6; DB 24; Length 8964;
 Matches 78; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

OY 72 ttgagtagtaaatattgtgctaagtaactattttagtagtgcagctgttatca 131
 DB 7681 ttaattagaatatttctgttaattatagaagcgttttggaggttttttagattga 7740
 OY 132 tatatcatcctcttctttagtcttctgtttaatggcagctacctctctgc 191
 DB 7741 tttttgtttagatctgttatattttaggtcttctgttattttagttttat 7800
 OY 192 atctagcagaatgagctgctgcagtt 218
 DB 7801 ttgtagttgtgtagtgcgttttagtt 7827

RESULT 11
 AAS61276
 ID AAS61276 standard; DNA; 8964 BP.
 AC AAS61276;
 XX
 DT 29-JAN-2002 (first entry)
 XX
 DE Human gene regulation-associated gene oligonucleotide #231.
 XX
 KW Human: Gene regulation-associated gene; severe combined immunodeficiency;
 KW cardiac damage; inflammatory response; Hemophilia; Werner syndrome;
 KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
 KW renal disease; Preeclampsia; cardiac allograft vascular disease;
 KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
 KW immunosuppressant; cardiac; antileukemic; coagulant; antileukemic;
 KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytosinetic;
 KW
 OS Homo sapiens.
 XX
 PN WO200177375-A2.
 XX
 PD 18-OCT-2001.
 XX
 PF 06-APR-2001; 2001WO-EP03968.
 XX
 PR 06-APR-2000; 2000DE-1019058.
 PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 XX

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: September 24, 2002, 14:45:19 ; Search time 3309.28 seconds
(without alignments)
2067.808 Million cell updates/sec

Title: US-09-617-174B-1_COPY_451_957

Perfect score: 507

Sequence: 1 tgagtaatcctaatacacagg.....ccaagaggcttgtagtaggag 507

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43.4	8.6	892	12 A2539346	A2539346 ENTGP83TR
2	41.8	8.2	1201	12 CNS016FX	AL106695 Drosophil
3	40.6	8.0	681	12 A0421326	A0421326 RPCI-11-1
4	40.2	7.9	1042	12 CNS024NB	AL180992 Tetraodon
5	40	7.9	1006	12 CNS03ESG	AL240793 Tetraodon
6	39.8	7.9	774	9 AL546807	AL546807 AL546807
7	39.6	7.8	457	10 BE943660	BE943660 UI-M-BH3-
8	39.6	7.8	494	9 A1848569	A1848569 UI-M-AM1-
9	39.6	7.8	1142	10 BM464476	BM464476 AGENCOURT
10	39.2	7.7	928	12 CNS00DKY	AL071865 Drosophil
11	39	7.7	850	12 A2673135	A2673135 ENTLI30TR
12	38.8	7.7	648	12 BH587629	BH587629 BOHEX88TF
13	38.6	7.6	365	9 A1605224	A1605224 ve78a05.x
14	38.6	7.6	388	12 B58764	B58764 CIT-RSP-201
15	38.6	7.6	419	10 BM163067	BM163067 EST565590
16	38.6	7.6	711	10 BF468105	BF468105 UI-M-CD0-
17	38.4	7.6	443	12 A0450690	A0450690 HS_5161_A

c 18	38.4	7.6	507	12 CNS01852	AL108896 Drosophil
19	38.4	7.6	726	10 BM170969	BM170969 EST573492
c 20	38.4	7.6	884	12 CNS00CXG	AL060063 Drosophil
c 21	38.4	7.6	1099	12 CNS01481	AL103836 Drosophil
22	38.4	7.6	1101	12 CNS0039G	AL063921 Drosophil
23	38.2	7.5	641	12 AG066397	AG066397 Pan trogl
24	38.2	7.5	907	12 CNS07COV	AL439445 T7 end of
25	37.8	7.5	535	12 AZ748844	AZ748844 RPCI-24-1
c 26	37.8	7.5	878	12 CNS0187R	AL108993 Drosophil
c 27	37.8	7.5	932	12 CNS0180Y	AL108748 Drosophil
c 28	37.8	7.5	935	12 CNS00J34	AL076343 Drosophil
c 29	37.8	7.5	1023	12 CNS00CVF	AL059889 Drosophil
c 30	37.8	7.5	1101	12 CNS008X3	AL052544 Drosophil
c 31	37.6	7.4	312	12 AZ821437	AZ821437 2M0094009
c 32	37.6	7.4	396	12 AQ683613	AQ683613 HS_5458_B
33	37.6	7.4	873	12 CNS024M5	AL180950 Tetraodon
c 34	37.4	7.4	797	10 BF253656	BF253656 HVSMEF000
c 35	37.4	7.4	1007	12 CNS01413	AL104181 Drosophil
c 36	37.4	7.4	1101	12 CNS0006J	AL062049 Drosophil
c 37	37.2	7.3	353	9 AW627276	AW627276 pa28a06.y
c 38	37.2	7.3	589	10 BE886609	BE886609 601509355
c 39	37.2	7.3	607	12 AG068424	AG068424 Pan trogl
c 40	37.2	7.3	614	9 AL655660	AL655660 AL655660
c 41	37.2	7.3	761	12 BH470789	BH470789 BOGYM20TR
c 42	37.2	7.3	767	12 CNS00AQX	AL055924 Drosophil
c 43	37.2	7.3	865	12 AZ675018	AZ675018 ENTHS57TR
44	37.2	7.3	889	12 AZ543817	AZ543817 ENFTS3TF
c 45	37	7.3	500	9 AU087167	AU087167 AU087167

ALIGNMENTS

RESULT 1
A2539346
LOCUS
DEFINITION ENTGP83TR Entamoeba histolytica Sheared DNA linear GSS 14-NOV-2000
genomic, DNA sequence.
ACCESSION A2539346
VERSION A2539346.1 GI:11145135
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica
REFERENCE Eukaryota; Entamoebidae; Entamoeba.
AUTHORS 1 (bases 1 to 892)
Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI:IMSS sheared DNA library
COMMENT Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjlloftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI-IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 15
High quality sequence stop: 847.

FEATURES

Location/Qualifiers
1..892
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHO51; Site:1; Bst.I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a

BASE COUNT	397 d	199 c	192 y	200 c	192 c	192 c
ORIGIN						

FEATURES	source	Class: BAC ends:	Location/Qualifiers
			1. .681
			/organism="Homo sapiens"
			/db_xref="GDB:7563969"
			/db_xref="taxon:9606"
			/clone="RPC1-11-167J10"
			/clone_lib="RPC1-11"
			/sex="Male"
			/cell_type="Lymphocytes"
			/note="vector: pBACE3.6; Site_1: EcoRI;
			Site_2: EcoRI;
			RPC11 Human Male BAC Library"

①
 ②
 ③
 ④

QY 138 cattctctcttttttgattgattgtctttctg 165
||||| ||| ||| ||| |||
Db 446 TATTCTCTTATTTCATTTCTCTGCTG 473

RESULT 6
AL546807 774 bp mRNA linear EST 16-FEB-2001
LOCUS AL546807 LTI_NFL006_PL2 Homo sapiens cDNA clone C50D1026Y122 5
DEFINITION prime, mRNA sequence.
ACCESSION AL546807
VERSION AL546807.1 GI:12880281
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 774)
AUTHORS Li W.B., Gruber C., Jessee J. and Polayes D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1. 774
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0D1026Y122"
/clone_lib="LTI_NFL006_PL2"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/
BASE COUNT 188 a 56 c 111 g 318 t 101 others
ORIGIN

Query Match 7.98; Score 39.8; DB 9; Length 774;
Best Local Similarity 37.28; Pred. No. 36; Indels 0; Gaps 0;
Matches 89; Conservative 34; Mismatches 116;
QY 72 ttgtgtagtaaatatttgcgtacgtacctaatttttagtgaggcactgtgtttataca 131
Db 66 WTTTWTATAGGTGCAATWTKAWTTWNGATTATTWTTTCTTTTGTAGCAATTAGA 125
QY 132 tatatccattctcttttttgattgattgtctttctgttaattgaaggcagctacctcttggc 191
Db 126 WAWATTATTATTCWTTTWTWAKGATTTTATTATTATAAAATTCGWTWTTAAATTTT 185
QY 192 atctagcagaatgagctgctgagttacacaaagaatgagagagagactctttt 251
Db 186 WTGCTCGTTTTAACCGTGTCTGTTKAGCWKTAGAGAKTWTWATGTYTTTTTTT 245
QY 252 tgccaccaacgtgtctgagaaattgtgtgtagtactatcatcacattattttt 310
Db 246 TTTTAAWAAKTTTWTWTWTTTWTWAAWTTTWTAKTAKYAAAYATTTTGTGKTK 304
RESULT 7
BE943660 457 bp mRNA linear EST 03-OCT-2000
LOCUS BE943660
DEFINITION UI-M-BH3-awf-g-11-0-UI.s1 NIH_BMAP_M_S4 Mus musculus cDNA clone
UI-M-BH3-awf-g-11-0-UI 3', mRNA sequence.
ACCESSION BE943660

VERSION BE943660.1 GI:10521419
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 457)
AUTHORS Bonaldo M.F., Lennon G. and Soares M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mEST@mail.nih.gov
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bona fide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the amygdala tissue cDNA library preparation: M.B. Soares Lab Clone distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements
Seq primer: M13 Forward
POLYA-Yes.

Location/Qualifiers
1. 457
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BH3-awf-g-11-0-UI"
/clone_lib="NIH_BMAP_M_S4"
/lab_host="NIH_BMAP_M_S4"
/note="Vector: p7T13D-Pac (Pharmacia) with a modified polylinker: Site 1: Not I; Site 2: Eco RI; The NIH_BMAP_M_S4 library is a subtracted library of a series, ultimately derived from a mixture of individually tagged normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hippocampus) after a series of subtractions to reduce the representation of cDNAs from which ESTs had already been generated. The following serially subtracted libraries were generated in this process: NIH_BMAP_M_S4, NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S1, NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library (NIH_BMAP_M_S4) was constructed as follows: PCR amplified cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with a pool of the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1 libraries in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the NIH_BMAP_M_S4 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
TAG_LIB=NIH_BMAP_M_S4
TAG_TISSUE=amygdala
TAG_SEQ=GTGAG"
117 a 103 c 87 g 150 t

BASE COUNT 117 a 103 c 87 g 150 t
ORIGIN


```

Query Match      7.8%; Score 39.6; DB 10; Length 457;
Best Local Similarity 66.3%; Pred. No. 41;
Matches 57; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 72 ttgtgtagtaaatatttctgaatagtagtaatttttagtaggcactgtgtttataca 131
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 15 TTTGCAAGTCAATATTAGCTAGTGAATTAATATTCTCTAGGAATGTTGGTTATCCA 74
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 132 tatatcattccctcttttttgattg 157
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 75 ACTACCTATTCTTTTACTTTG 100
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 8
LOCUS      A1848569          494 bp      mRNA      linear      EST 15-JUL-1999
DEFINITION UI-M-AM1-aga-b-02-0-UI.s2 NIH_BMAP_MAM_N Mus musculus cDNA clone
ACCESSION  A1848569
VERSION    A1848569.1 GI:5492475
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus

REFERENCE  1 (bases 1 to 494)
AUTHORS   Bonaldo, M.F., Lennon, G. and Soares, M.B.
TITLE     Normalization and subtraction: two approaches to facilitate gene
JOURNAL   Genome Res. 6 (9), 791-806 (1996)
MEDLINE   9704477
COMMENT   Contact: Chin, H
          National Institute of Mental Health
          6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
          20892-9643, USA
          Tel: 301 443 1706
          Fax: 301 443 9890
          Email: mEST@mail.nih.gov
          The sequence contained an oligo-dT track that was present in the
          oligonucleotide that was used to prime the synthesis of first
          strand cDNA and therefore this may represent a bonafide poly A
          tail. The sequence tag present in the cDNA between the NotI site
          and the oligo-dT track served to verify it as a clone from the
          normalized amygdala library cDNA Library Preparation: M.B. Soares
          Lab Clone distribution: NIH BMAP cDNA clones will be made available
          by the means that is soon to be determined. When NIH determines the
          means for distribution of the BMAP cDNA clones, this record will be
          updated accordingly when that means is determined.
          Seq primer: M13 Forward
          POLYA=Yes.

          Location/Qualifiers
          1..494
             /organism="Mus musculus"
             /strain="C57BL/6J"
             /db_xref="taxon:10090"
             /clone="UI-M-AM1-aga-b-02-0-UI"
             /clone_lib="NIH_BMAP_MAM_N"
             /dev_stage="27-32 days"
             /lab_host="DH10B (Life Technologies)"
             /note="Vector: pT73D-Pac (Pharmacia) with a modified
             polylinker. Site.1: Not I; Site.2: Eco RI; The
             NIH_BMAP_MAM_N library is a normalized library constructed
             from mouse amygdala. The tag is a string of 5 nucleotides
             present between the Not I site and the oligo-dT track.
             The library was constructed as described by Bonaldo,
             Lennon and Soares, Genome Research 6: 791-806, 1996.
             Tissue provided by Ms. Annie Novakovich, Zivic-Miller
             Laboratories.
             TAG_LIB=NIH_BMAP_MAM_N
             TAG_TISSUE=amygdala
             TAG_SEQ=GTGAG"

BASE COUNT      126 a 118 c 91 g 159 t

```

ORIGIN

```

Query Match      7.8%; Score 39.6; DB 9; Length 494;
Best Local Similarity 56.3%; Pred. No. 41;
Matches 57; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 72 ttgtgtagtaaatatttctgaatagtagtaatttttagtaggcactgtgtttataca 131
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 15 TTTGCAAGTCAATATTAGCTAGTGAATTAATATTCTCTAGGAATGTTGGTTATCCA 74
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 132 tatatcattccctcttttttgattg 157
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 75 ACTACCTATTCTTTTACTTTG 100
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 9
LOCUS      BM464476          1142 bp      mRNA      linear      EST 05-FEB-2002
DEFINITION AGENCOURT_6438833 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:553563
ACCESSION  BM464476
VERSION    BM464476
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens

REFERENCE  1 (bases 1 to 1142)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: ATCC
          cDNA Library Preparation: Life Technologies, Inc.
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLAM12223 row: n column: 12
          High quality sequence stop: 630.

          Location/Qualifiers
          1..1142
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="IMAGE:553563"
             /clone_lib="NIH_MGC_71"
             /tissue_type="leiomyosarcoma"
             /lab_host="DH10B (phage-resistant)"
             /note="Organ: uterus; Vector: pCMV-SPORT6; Site.1: NotI;
             Site.2: SalI; Cloned unidirectionally. Primer: Oligo df.
             Average insert size 2.1 kb."
             378 a 245 c 224 g 259 t 36 others

BASE COUNT      378 a 245 c 224 g 259 t 36 others
ORIGIN

Query Match      7.8%; Score 39.6; DB 10; Length 1142;
Best Local Similarity 48.6%; Pred. No. 39;
Matches 90; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 105 tttttagtaggcactgttttatcatatcatcattcttttttttttttttttttct 164
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 645 TTTTNNNNNNNTTTTNTNTNTTTTNTNTNTTTTNTNTNTTTTNTNTNTTTTNTNT 586
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 165 gtttaaggcagctactctctcttgccatctagcagaatgagctgctgcagtttacacaa 224
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 585 TTTTGTGCACATATACAGTACTTATTGTACAAATAAGGAATGGGGAAGGGAATG 526
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 225 aaagaatggagatcagagtcactttttgtgccaccaacgtgtctgagaaattttagtgtt 284
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 525 AAAGAAATAGAGAAACTATACGGTAGTAGTACGTAGGTGTGTGGAAACAAATTCAGTTT 466
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```



```

Db 124 AATTATC 130
| | | |
RESULT 12
BH587629/c
LOCUS BH587629.1 G1:17840087
DEFINITION BOHE Brassica oleracea genomic clone BOHEX88, DNA
sequence.
ACCESSION BH587629
VERSION BH587629.1
KEYWORDS GSS.
SOURCE Brassica oleracea.
ORGANISM Brassica oleracea
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
TITLE Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
JOURNAL 1 (bases 1 to 648)
COMMENT Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished (2001)
Other GSSs: BOHEX88TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
FEATURES
Location/Qualifiers
source
1..648
/organism="Brassica oleracea"
/strain="TOL000DH3"
/db_xref="taxon:3712"
/clone="BOHEX88"
/clone_lib="BOHE"
/note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT 254 a 71 c 60 g 263 t
ORIGIN
Query Match 7.7%; Score 38.8; DB 12; Length 648;
Best Local Similarity 50.5%; Pred. No. 62;
Matches 94; Conservative 0; Mismatches 92; Indels 0; Gaps 0;
QY 52 taacaataatctctaccatttggttagtaataattttgctaatagtagcctaatatttttag 111
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 622 TTAATAACATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 563
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 112 gtaggcactggtttatcacatatatccattctctctttttgattgctctttctgtttaat 171
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 562 TTAGGCACAGAGATAATACTATAAATTTTAAAGCATTTATGATGATTTCTTGTATTTTAT 503
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 172 gggcagctacctctctctgcatctgcagcagatgagctgctgcagtttccacaaaagaat 231
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 502 GGATATTTCCGAATATTTACGGATGTTTAAAGATATCTTATCATCTTATACATAACAAT 443
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 232 ggagat 237
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 442 AAAAAT 437
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
RESULT 13
AI605224/c
LOCUS AI605224.1 G1:832304
DEFINITION ve78a05_x1 Soares_mammary_gland_NBMGM Mus musculus cDNA clone
365 bp mRNA linear EST 21-APR-1999
ACCESSION AI605224
VERSION AI605224
KEYWORDS IMAGE:832304 3', mRNA sequence.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 388)
house mouse.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 365)
JOURNAL NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
COMMENT National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:492520
This clone was previously sequenced on the 5' end only, this new
data is from the 3' end
High quality sequence stop: 357.
FEATURES
Location/Qualifiers
source
1..365
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:832304"
/clone_lib="Soares_mammary_gland_NBMGM"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland; Vector: p773D-Pac (Pharmacia
) with a modified polylinker; Site_1: Not I; Site_2: Eco
RI; 1st strand cDNA was primed with a Not I - Oligo(dT)
primer [5'
TGTTCACCAATCTGAAGTGGAGCGCGCGAATGCTTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified p773 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."
BASE COUNT 136 a 72 c 53 g 103 t 1 others
ORIGIN
Query Match 7.6%; Score 38.6; DB 9; Length 365;
Best Local Similarity 53.7%; Pred. No. 71;
Matches 80; Conservative 0; Mismatches 69; Indels 0; Gaps 0;
QY 47 ccattaaacaatatctctaccatttggttagtaataattttgctaatagtagcctaat 106
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 273 CCAATAAATAATGATAAACTTTCTTTTAGGCTCTTTATTGTTGGCTGTCTATGCTAGATT 214
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 107 tttaggtaggcactggtttatcacatatatccattctctctttttgattgctcttctgt 166
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 213 ATATAGTATTAACCTAAATAACTACCTGTCTCCATCTCTTTATTATTGTTGTTGTTGT 154
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 167 ttaatggggcagctacctctctctggcatct 195
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 153 TTGTTGTTGTTGTTTCAAGACAGCGTTTCT 125
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
RESULT 14
B58764/c
LOCUS B58764.1 G1:2613482
DEFINITION CIT-HSP-2013020.TF CIT-HSP Homo sapiens genomic clone 2013020, DNA
sequence.
ACCESSION B58764
VERSION B58764.1
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 388)

```

Email: carltonetigr.org
For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ARCC
<http://www.malaria.mr4.org/mr4pages/index.html>
Seq primer: ADF.

FEATURES	SOURCE
----------	--------

```
l. 419
/organism="plasmodium yoelii yoelii"
/strain="17XL"
/db_xref="/db_taxon:73239"
/clone="PYCIA48"
/clone_lib="pyBS"
/dev stages="Asexual blood stages"
```

/lab_host="E. coli XL-1 Blue"/note="PAD-GAL4"; At 20-25% parasitemia, blood was collected from BALB/cByJ mice infected with Py1/XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidinium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with xhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybriZAP arms directionally using EcoRI-xhoI cleaved arms. After packaging, the phagemid vector (PAD-GAL4) was excised from the HybriZAP vector and plasmid DNA isolated."

BASE COUNT	226 a	37 c	49 g	107 t
ORIGIN				

Query Match	7.6%	Score 38.6;	DB 10;	Length 419;
Best Local Similarity	51.4%	Pred. No. 71;		
Matches	89:	Conservative	0:	Mismatches 84;
				Indels 0;
				Gaps 0;

[illegible]

Search completed: September 24, 2002, 14:45:21
Job time: 16656 sec

Email: carltonet@tigr.org
For more clone info, please contact the Malaria Research and Reference Reagent Resource Center, ATCC
<http://www.malaria.mr4.org/mr4pages/index.html>
See primer: ADF.

FEATURES	SOURCE
1. The first 100 words of the text are used to generate a vector.	1. The first 100 words of the text are used to generate a vector.
2. The vector is then used to predict the next word in the sequence.	2. The vector is then used to predict the next word in the sequence.
3. The process is repeated for the next 100 words, and so on.	3. The process is repeated for the next 100 words, and so on.
4. The final output is a list of predicted words.	4. The final output is a list of predicted words.

```
l. 419
/organism="plasmodium yoelii yoelii"
/strain="17XL"
/db_xref="/db_taxon:73239"
/clone="PYCIA48"
/clone_lib="pyBS"
/dev stages="Asexual blood stages"
```

/lab_host="E. coli XL-1 Blue"/note="PAD-GAL4"; At 20-25% parasitemia, blood was collected from BALB/cByJ mice infected with Py1/XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidinium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with xhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybriZAP arms directionally using EcoRI-xhoI cleaved arms. After packaging, the phagemid vector (PAD-GAL4) was excised from the HybriZAP vector and plasmid DNA isolated."

BASE COUNT	226 a	37 c	49 g	107 t
ORIGIN				

Query Match	7.6%	Score 38.6;	DB 10;	Length 419;
Best Local Similarity	51.4%	Pred. No. 71;		
Matches	89;	Conservative	0;	Mismatches 84;
		Indels	0;	Gaps 0;

[illegible]

Search completed: September 24, 2002, 14:45:21
Job time: 16656 sec

RESULT	15
BMBM163067/c	
LOCUS	419 bp mRNA linear EST 04-DEC-2001
DEFINITION	EST565590 PyBS Plasmodium yoelii cDNA clone PYCLA48 5' end,
ACCESSION	BM163067
VERSION	BM163067.1 GI:17308748
KEYWORDS	mRNA sequence.
SOURCE	Plasmodium yoelii yoelii.
ORGANISM	Plasmodium yoelii yoelii.
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS	1 (bases 1 to 419)
TITLE	Fraser,C.M. and Carucci,D.J.
JOURNAL	Plasmodium yoelii EST project at TIGR
COMMENT	Unpublished (2001) Contact: Jane Carlton Parasite Genomics Group The Institute for Genomic Research 9712 Medical Center Drive, Rockville, MD 20850, USA Tel: 301-530-9319 Fax: 301-838-0208

•
•
•
•

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: September 24, 2002, 15:52:52 ; Search time 3102.12 seconds
(without alignments)
296.819 Million cell updates/sec

Title: US-09-617-174B-1_COPY_914_957

Perfect score: 44

Sequence: 1 tctcttcaattgtgacaa.....ccaagaggcttagtaggag 44

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_em.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pr.*

9: gb_ro.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	ID	Description
------------	-------------	--------	----	-------------

ALIGNMENTS

RESULT 1

AC090307 AC090307 170269 bp DNA linear PRI 30-DEC-2001
LOCUS Homo sapiens chromosome 18, clone RP11-851B10, complete sequence.

DEFINITION AC090307

ACCESSION AC090307

VERSION AC090307.7 GI:16041402

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 170269)

2 (bases 1 to 170269)

Unpublished

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Boguslavsky, L., Bouckgeert, B., Brown, A.,
Camarata, J., Campopiano, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J.,
Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,

Jones, C., Karatas, A., LaRocque, K., Lamazares, R., Landers, T.,
 Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P.,
 Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K.,
 McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mienga, V.,
 Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H.,
 O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
 Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R.,
 Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M.,
 Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P.,
 Sougnez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
 Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A.,
 Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J.,
 Zembek, L., Zimmer, A. and Zody, M.

TITLE

JOURNAL
 Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE

AUTHORS
 3 (bases 1 to 170269)
 Birren, B., Linton, L., Nusbaum, C., Landers, E., Ali, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,
 Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
 Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A.,
 Cooke, P., DeArelano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,
 Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
 Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,
 Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
 Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
 McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J.,
 Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,
 Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R.,
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

JOURNAL
 Submitted (30-DEC-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

On Oct 11, 2001 this sequence version replaced gi:15290867.
 All repeats were identified using RepeatMasker:
 Smit, A. F. A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L12576

Center clone name: 851_B_10

FEATURES

source

Location/Qualifiers

1. .170269

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="18"

/map="18"

/clone="RP11-851B10"

/clone_lib="RPC1-11 Human Male BAC"

/complement(2. .177)

/rpt_family="L1LMC5"

/complement(189. 489)

/rpt_family="AluSx"

/complement(708. .803)

/rpt_family="MIR"

/complement(1616. .1904)

/rpt_family="AluJo"

/complement(2815. .3120)

repeat_region
 /rpt_family="AluSx"
 3218. .3365
 /rpt_family="MIR"
 3565. .3590
 /rpt_family="(TG)n"
 complement(3782. .4028)
 /rpt_family="MIR"
 4056. .4177
 /rpt_family="MER5A"
 4601. .5620
 /rpt_family="HSMAR2"
 5621. .5643
 /rpt_family="(CAA)n"
 5644. .5866
 /rpt_family="HSMAR2"
 6184. .6475
 /rpt_family="L1MA2"
 6677. .6837
 /rpt_family="(TA)n"
 6852. .7018
 /rpt_family="(TA)n"
 6994. .7122
 /rpt_family="(CATATA)n"
 7146. .7332
 /rpt_family="(CATATA)n"
 7371. .7526
 /rpt_family="(TA)n"
 complement(7527. .7835)
 /rpt_family="AluSx"
 7936. .8080
 /rpt_family="MIR"
 8697. .8927
 /rpt_family="AluSx"
 complement(9036. .9288)
 /rpt_family="AluSx"
 9873. .9934
 /rpt_family="(GA)n"
 10340. .10367
 /rpt_family="AT-rich"
 10368. .10686
 /rpt_family="AluY"
 10687. .10711
 /rpt_family="AT-rich"
 complement(10870. .11043)
 /rpt_family="L2"
 11348. .11375
 /rpt_family="(T)n"
 12359. .12449
 /rpt_family="L3"
 complement(12634. .12890)
 /rpt_family="MIR"
 complement(13436. .13725)
 /rpt_family="AluJb"
 14244. .14297
 /rpt_family="(CGGG)n"
 complement(15188. .15350)
 /rpt_family="MIR3"
 16056. .16114
 /rpt_family="MIR3"
 complement(16129. .16429)
 /rpt_family="AluSx"
 16532. .16646
 /rpt_family="MER94"
 16855. .17145
 /rpt_family="L2"
 17377. .17572
 /rpt_family="MIR3"
 complement(20036. .20330)
 /rpt_family="Charliel"
 complement(20333. .20486)
 /rpt_family="Charliel"
 20625. .20691
 /rpt_family="L2"

repeat_region

```

Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zlmer,A. and Zody,M.
Direct Submission
Submitted (07-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Dec 13, 2001 this sequence version replaced gl:16949388.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L9215
Center clone name: 635_N_19
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 114304: contig of 114304 bp in length
* 114305 114404: gap of 100 bp
* 114405 173131: contig of 58727 bp in length.
Location/Qualifiers
1..173131
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-635N19"
/clone_lib="RPCI-11 Human Male BAC"
51584 a 35315 c 35660 g 50470 t 101 others
BASE COUNT
ORIGIN

Query Match 96.4%; Score 42.4; DB 2; Length 173131;
Best Local Similarity 97.7%; Pred. No. 1.2e-07;
Matches 43; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 tctcttcatttggaagcgtgcgaagaggcttgtagtaggag 44
+
|||||
Db 51903 TTCTCTTCAATTGGCAAGCTGCCAAGAGCGCTTGAGTAGGAG 51860

RESULT 3
AP001404/c
LOCUS
DEFINITION Homo sapiens chromosome 18 clone RP11-851B10 map 18q22, WORKING
DRAFT SEQUENCE, 11 unordered pieces.
ACCESSION AP001404
VERSION AP001404.3 GI:9188495
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens DNA, clone:RP11-851B10.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 173836)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 173,836 genomic DNA of 18q22
TITLE

```


Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
 Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

TITLE JOURNAL

Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 3, 2000 this sequence version replaced gi:6446894.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L1590

Center clone name: L117_P_15

* NOTE: This record contains 240 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1 695: contig of 695 bp in length
 696 795: gap of 100 bp
 796 1513: contig of 718 bp in length
 1514 1613: gap of 100 bp
 1614 2337: contig of 724 bp in length
 2338 2437: gap of 100 bp
 2438 3161: contig of 724 bp in length
 3162 3261: gap of 100 bp
 3262 3966: contig of 705 bp in length
 3967 4066: gap of 100 bp
 4067 4772: contig of 706 bp in length
 4773 4872: gap of 100 bp
 4873 5564: contig of 692 bp in length
 5565 5664: gap of 100 bp
 5665 6397: contig of 733 bp in length
 6398 6497: gap of 100 bp
 6498 7202: contig of 705 bp in length
 7203 7302: gap of 100 bp
 7303 8028: contig of 726 bp in length
 8029 8128: gap of 100 bp
 8129 8803: contig of 675 bp in length
 8804 8903: gap of 100 bp
 8904 9613: contig of 710 bp in length
 9614 9713: gap of 100 bp
 9714 10409: contig of 696 bp in length
 10410 10509: gap of 100 bp
 10510 11237: contig of 728 bp in length
 11238 11337: gap of 100 bp
 11338 12043: contig of 706 bp in length
 12044 12143: gap of 100 bp
 12144 12867: contig of 724 bp in length
 12868 12967: gap of 100 bp
 12968 13660: contig of 693 bp in length
 13661 13760: gap of 100 bp
 13761 14489: contig of 729 bp in length
 14490 14589: gap of 100 bp
 14590 15314: contig of 725 bp in length
 15315 15414: gap of 100 bp
 15415 16129: contig of 715 bp in length
 16130 16229: gap of 100 bp
 16230 16946: contig of 717 bp in length
 16947 17046: gap of 100 bp
 17047 17769: contig of 723 bp in length

17770 17869: gap of 100 bp
 17870 18574: contig of 705 bp in length
 18575 18674: gap of 100 bp
 18675 19379: contig of 705 bp in length
 19380 19479: gap of 100 bp
 19480 20186: contig of 707 bp in length
 20187 20236: gap of 100 bp
 20287 20980: contig of 694 bp in length
 20981 21080: gap of 100 bp
 21081 21779: contig of 699 bp in length
 21780 21879: gap of 100 bp
 21880 22594: contig of 715 bp in length
 22595 22694: gap of 100 bp
 22695 23399: contig of 705 bp in length
 23400 23499: gap of 100 bp
 23500 24217: contig of 718 bp in length
 24218 24317: gap of 100 bp
 24318 25026: contig of 709 bp in length
 25027 25126: gap of 100 bp
 25127 25836: contig of 710 bp in length
 25837 25936: gap of 100 bp
 25937 26637: contig of 701 bp in length
 26638 26737: gap of 100 bp
 26738 27446: contig of 709 bp in length
 27447 27546: gap of 100 bp
 27547 28274: contig of 728 bp in length
 28275 28374: gap of 100 bp
 28375 29095: contig of 721 bp in length
 29096 29155: gap of 100 bp
 29156 29917: contig of 722 bp in length
 29918 30017: gap of 100 bp
 30018 30724: contig of 707 bp in length
 30725 30824: gap of 100 bp
 30825 31529: contig of 705 bp in length
 31530 31629: gap of 100 bp
 31630 32324: contig of 695 bp in length
 32325 32424: gap of 100 bp
 32425 33116: contig of 692 bp in length
 33117 33216: gap of 100 bp
 33217 33936: contig of 720 bp in length
 33937 34036: gap of 100 bp
 34037 34750: contig of 714 bp in length
 34751 34850: gap of 100 bp
 34851 35578: contig of 728 bp in length
 35579 35678: gap of 100 bp
 35679 36393: contig of 715 bp in length
 36394 36493: gap of 100 bp
 36494 37203: contig of 710 bp in length
 37204 37303: gap of 100 bp
 37304 37965: contig of 662 bp in length
 37966 38065: gap of 100 bp
 38066 33781: contig of 716 bp in length
 38782 38881: gap of 100 bp
 38882 39590: contig of 709 bp in length
 39591 39690: gap of 100 bp
 39691 40403: contig of 713 bp in length
 40404 40503: gap of 100 bp
 40504 41238: contig of 735 bp in length
 41239 41338: gap of 100 bp
 41339 42064: contig of 726 bp in length
 42065 42164: gap of 100 bp
 42165 42880: contig of 716 bp in length
 42881 42980: gap of 100 bp
 42981 43687: contig of 707 bp in length
 43688 43787: gap of 100 bp
 43788 44504: contig of 717 bp in length
 44505 44604: gap of 100 bp
 44605 45307: contig of 703 bp in length
 45308 45407: gap of 100 bp
 45408 46106: contig of 699 bp in length
 46107 46206: gap of 100 bp
 46207 46909: contig of 703 bp in length
 46910 47009: gap of 100 bp

All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: W1BR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

Project Information
Center project name: L1590
Center clone name: 1117_D_15

* NOTE: This record contains 240 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1 695: contig of 695 bp in length
* 696 795: gap of 100 bp
* 796 1513: contig of 718 bp in length
* 1514 1613: gap of 100 bp
* 1614 2337: contig of 724 bp in length
* 2338 2437: gap of 100 bp
* 2438 3161: contig of 724 bp in length
* 3162 3261: gap of 100 bp
* 3262 3966: contig of 705 bp in length
* 3967 4066: gap of 100 bp
* 4067 4772: contig of 706 bp in length
* 4773 4872: gap of 100 bp
* 4873 5564: contig of 692 bp in length
* 5565 5664: gap of 100 bp
* 5665 6397: contig of 733 bp in length
* 6398 6497: gap of 100 bp
* 6498 7202: contig of 705 bp in length
* 7203 7302: gap of 100 bp
* 7303 8028: contig of 726 bp in length
* 8029 8128: gap of 100 bp
* 8129 8803: contig of 675 bp in length
* 8804 8903: gap of 100 bp
* 8904 9613: contig of 710 bp in length
* 9614 9713: gap of 100 bp
* 9714 10409: contig of 696 bp in length
* 10410 10509: gap of 100 bp
* 10510 11237: contig of 728 bp in length
* 11238 11337: gap of 100 bp
* 11338 12043: contig of 706 bp in length
* 12044 12143: gap of 100 bp
* 12144 12867: contig of 724 bp in length
* 12868 12967: gap of 100 bp
* 12967 13660: contig of 693 bp in length
* 13661 13760: gap of 100 bp
* 13761 14489: contig of 729 bp in length
* 14490 14589: gap of 100 bp
* 14590 15314: contig of 725 bp in length
* 15315 15414: gap of 100 bp
* 15415 16129: contig of 715 bp in length
* 16130 16229: gap of 100 bp
* 16230 16946: contig of 717 bp in length
* 16947 17046: gap of 100 bp
* 17047 17769: contig of 723 bp in length
* 17770 17869: gap of 100 bp
* 17870 18574: contig of 705 bp in length
* 18575 18674: gap of 100 bp
* 18675 19379: contig of 705 bp in length
* 19380 19479: gap of 100 bp
* 19480 20186: contig of 707 bp in length
* 20187 20286: gap of 100 bp

Query Match 96.4%; Score 42.4; DB 2; Length 214212;
Best Local Similarity 97.7%; Pred. No. 1.2e-07;
Matches 43; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 tctcttcaattgtgacagctcccaaggcttgagtaggag 44
Db 57083 TTTCTTTCAATTGTGACAGCTCCCAAGGCTTGAGTAGGAG 57126

RESULT 5
AC015793/c
LOCUS AC015793
DEFINITION Homo sapiens chromosome 18 clone RP11-1117D15 map 18, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC015793
VERSION AC015793.2 GI:7144769
KEYWORDS HTG: HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 214212)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Homo sapiens chromosome 18, clone RP11-1117D15
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 214212)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Boukhgalter, B.,
Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,
Cooke, P., Dearellano, K., Dewar, K., Domino, M., Donellan, L., Doyle, M.,
Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Lehoczy, J., Lieu, C., Locke, K., MacDonald, P., Marquis, N.,
McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J.,
Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye-W.J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Mar 3, 2000 this sequence version replaced gi:6446894.

20287	20980: contig of 694 bp	in length
20981	21080: gap of 100 bp	
21081	21779: contig of 699 bp	in length
21780	21879: gap of 100 bp	
21880	22594: contig of 715 bp	in length
22595	22694: gap of 100 bp	
22695	23399: contig of 705 bp	in length
23400	23499: gap of 100 bp	
23500	24217: contig of 718 bp	in length
24218	24317: gap of 100 bp	
24318	25026: contig of 709 bp	in length
25027	25126: gap of 100 bp	
25127	25836: contig of 710 bp	in length
25937	25936: gap of 100 bp	
25937	26637: contig of 701 bp	in length
26638	26737: gap of 100 bp	
26738	27446: contig of 709 bp	in length
27447	27546: gap of 100 bp	
27547	28274: contig of 728 bp	in length
28275	28374: gap of 100 bp	
28375	29095: contig of 721 bp	in length
29096	29195: gap of 100 bp	
29196	29917: contig of 722 bp	in length
29918	30017: gap of 100 bp	
30018	30724: contig of 707 bp	in length
30725	30824: gap of 100 bp	
30825	31529: contig of 705 bp	in length
31530	31629: gap of 100 bp	
31630	32324: contig of 695 bp	in length
32325	32424: gap of 100 bp	
32425	33116: contig of 692 bp	in length
33117	33216: gap of 100 bp	
33217	33936: contig of 720 bp	in length
33937	34036: gap of 100 bp	
34037	34750: contig of 714 bp	in length
34751	34850: gap of 100 bp	
34851	35578: contig of 728 bp	in length
35579	35678: gap of 100 bp	
35679	36393: contig of 715 bp	in length
36394	36493: gap of 100 bp	
36494	37203: contig of 710 bp	in length
37204	37303: gap of 100 bp	
37304	37965: contig of 662 bp	in length
37966	38065: gap of 100 bp	
38066	38781: contig of 716 bp	in length
38782	38881: gap of 100 bp	
38882	39590: contig of 709 bp	in length
39591	39690: gap of 100 bp	
39691	40403: contig of 713 bp	in length
40404	40503: gap of 100 bp	
40504	41238: contig of 735 bp	in length
41239	41338: gap of 100 bp	
41339	42064: contig of 726 bp	in length
42065	42164: gap of 100 bp	
42165	42880: contig of 716 bp	in length
42881	42980: gap of 100 bp	
42981	43687: contig of 707 bp	in length
43688	43787: gap of 100 bp	
43788	44504: contig of 717 bp	in length
44505	44604: gap of 100 bp	
44605	45307: contig of 703 bp	in length
45308	45407: gap of 100 bp	
45408	46106: contig of 699 bp	in length
46107	46206: gap of 100 bp	
46207	46909: contig of 703 bp	in length
46910	47009: gap of 100 bp	
47010	47696: contig of 687 bp	in length
47697	47796: gap of 100 bp	
47797	48531: contig of 735 bp	in length
48532	48631: gap of 100 bp	
48632	49341: contig of 710 bp	in length
49342	49441: gap of 100 bp	in length
49442	50162: contig of 721 bp	in length

Query Match	95.4%	Score 42.4;	DB 2;	Length 214212;
Best Local Similarity	97.7%	Pred. No. 1.2e-07;		
Matches	43;	Conservative	0;	Mismatches 1; Indels 0; Gaps 0;
Qy	1	tctcttcaattggaagaagctgccaaagagcgttgagtaggag	44	
Db	53840	TTTTTTCATTTGGACACAGCTGCCAAGAGGCTTGAGTAGGAG	53797	
RESULT	6			
AX251235				
LOCUS	AX251235	Sequence 203 from Patent WO0168912.	12426 bp	DNA linear PAT 05-OCT-2001
DEFINITION	AX251235			
ACCESSION	AX251235			
VERSION	AX251235.1	GI:15984658		
KEYWORDS		synthetic construct		
SOURCE		synthetic construct		
ORGANISM		artificial sequence		
REFERENCE		1 (bases 1 to 12426)		
AUTHORS		Olek.A., Piepenbrock,C. and Berlin,K.		
TITLE		Diagnosis of diseases associated with tumor suppressor genes and oncogenes		
JOURNAL		Patent: WO 0168312-A 203 20-SEP-2001;		
FEATURES		Epigenomics AG (DE)		
source		Location/Qualifiers		
		1. .12426		
		/organism="synthetic construct"		
		/db_xref="taxon:32630"		
		/note="chemically treated genomic DNA (Homo sapiens)"		
BASE COUNT	3527 a	137 c	2591 g	6171 t
ORIGIN				
Query Match	70.9%	Score 31.2;	DB 6;	Length 12426;
Best Local Similarity	82.8%	Pred. No. 0.0093;		
Matches	36;	Conservative	0;	Mismatches 8; Indels 0; Gaps 0;
Qy	1	tctcttcaattggaagaagctgccaaagagcgttgagtaggag	44	
Db	4834	TTTTTTTAAATGTGGAT/AAGTTGTTAAGAGGTTTGACTAGGAG	4877	
RESULT	7			
AC025047/c				
LOCUS	AC025047	Mus musculus chromosome 4 clone RP23-5618 map 4, *** SEQUENCING IN	177842 bp	DNA linear HTG 01-FEB-2002
DEFINITION	AC025047			
ACCESSION	AC025047	PROGRESS ***; 3 ordered pieces.		
VERSION	AC025047.11	GI:18464142		
KEYWORDS		HTG; HTGS PHASE2; HTGS FULLTOP; HTGS ACTIVEFIN.		


```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 139476)
TITLE DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Direct Submission
UNPUBLISHED
REFERENCE
AUTHORS 2 (bases 1 to 139476)
TITLE DOE Joint Genome Institute.
JOURNAL Direct Submission
UNPUBLISHED
REFERENCE
AUTHORS 3 (bases 1 to 139476)
TITLE DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Direct Submission
UNPUBLISHED
REFERENCE
AUTHORS 4 (bases 1 to 139476)
TITLE DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Direct Submission
UNPUBLISHED
COMMENT On Oct 3, 2001 this sequence version replaced gi:10312245.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.4% of Sequence;
Estimated Total Number of Errors is 0.8.
STS Content:
SHGC-7426 GI4255
SHGC-36388 G30134
SHGC-16052 G15353.
FEATURES
source
1. 139476
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-201719"
BASE COUNT 42042 a 26255 c 27397 g 43782 t
ORIGIN
Query Match 56.4%; Score 24.8; DB 9; Length 139476;
Best Local Similarity 72.7%; Pred. No. 6.5;
Matches 32; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 1 tctcttcaattgtgacaaagctgccaaagagcttgagtaggag 44
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 122755 TCTCTCTCAATTTGGACACACGCGCGCTTACTTTAGTTGAAG 122712

RESULT 10
AC008455 141597 bp DNA linear HTG 19-APR-2001
LOCUS Homo sapiens chromosome 5 clone CTC-34711, WORKING DRAFT SEQUENCE,
DEFINITION 9 ordered pieces.
ACCESSION AC008455
VERSION AC008455.7 GI:13676949
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 141597)
TITLE DOE Joint Genome Institute.
JOURNAL Direct Submission
UNPUBLISHED
REFERENCE
AUTHORS 2 (bases 1 to 141597)
TITLE DOE Joint Genome Institute.
JOURNAL Direct Submission
UNPUBLISHED
COMMENT On Apr 19, 2001 this sequence version replaced gi:10305080.

```

```

-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 318619
Center clone name: CIT-HSPC_34711
-----
Summary Statistics
Consensus quality: 128407 bases at least Q40
Consensus quality: 138430 bases at least Q30
Consensus quality: 139955 bases at least Q20
Estimated insert size: 140000; pulse field gel estimation
Estimated insert size: 140797; sum-of-ctngs estimation
Quality coverage: 5.6 in Q20 bases; pulse field gel estimation
Quality coverage: 5.57 in Q20 bases; sum-of-ctngs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 26634: contig of 26634 bp in length
* 26635 26734: gap of unknown length
* 26735 34856: contig of 8122 bp in length
* 34857 34956: gap of unknown length
* 34957 61826: contig of 26870 bp in length
* 61827 61926: gap of unknown length
* 61927 65313: contig of 3387 bp in length
* 65314 65413: gap of unknown length
* 65414 67577: contig of 2164 bp in length
* 67578 67677: gap of unknown length
* 67678 100223: contig of 32546 bp in length
* 100224 100323: gap of unknown length
* 100324 125153: contig of 24830 bp in length
* 125154 125253: gap of unknown length
* 125254 133618: contig of 8365 bp in length
* 133619 133718: gap of unknown length
* 133719 141597: contig of 7879 bp in length.
FEATURES
Location/Qualifiers
1. 141597
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTC-34711"
/clone_lib="CalTech human BAC library C"
BASE COUNT 41447 a 26594 c 26971 g 45784 t 801 others
ORIGIN
Query Match 56.4%; Score 24.8; DB 2; Length 141597;
Best Local Similarity 72.7%; Pred. No. 6.5;
Matches 32; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 1 tctcttcaattgtgacaaagctgccaaagagcttgagtaggag 44
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 135935 TCTCTCTCAATTTGGACACACGCGCGCTTACTTTAGTTGAAG 135978

RESULT 11
AC016619/c 155459 bp DNA linear HTG 19-APR-2001
LOCUS Homo sapiens chromosome 5 clone CTD-2245F16, WORKING DRAFT
DEFINITION SEQUENCE, 6 ordered pieces.
ACCESSION AC016619
VERSION AC016619.5 GI:7711578
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens

```


Stantropop, S., Thomas, J.W., Thomas, P.J., Touchman, J.W.,
Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Zhang, L.-H. and Green, E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 90309)
Green, E.D.
Direct Submission
Submitted (27-OCT-2001) NIH Intramural Sequencing Center, 8717
Grovermont Circle, Gaithersburg, MD 20877, USA
On Nov 30, 2001 this sequence version replaced gi:16506397.
----- Genom Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: <http://www.nisc.nih.gov>
Contact: nisc.mouse@nih.gov
----- Project Information
Center project name: arp
Center clone name: 205P11
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 87445 bases at least Q40
Consensus quality: 88132 bases at least Q30
Consensus quality: 88439 bases at least Q20
Insert size: 89000; agarose-fp
Insert size: 89609; sum-of-contigs
Quality coverage: 10.67x in Q20 bases; agarose-fp
Quality coverage: 10.60x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 2330: contig of 2330 bp in length
* 2331
2330: gap of unknown length
* 2431
5020: contig of 2590 bp in length
* 5021
5120: gap of unknown length
* 5121
7685: contig of 2565 bp in length
* 7686
7785: gap of unknown length
* 7786
13195: contig of 5410 bp in length
* 13196
13295: gap of unknown length
* 13296
20253: contig of 6958 bp in length
* 20254
20353: gap of unknown length
* 20354
26321: contig of 5968 bp in length
* 26322
26421: gap of unknown length
* 26422
51911: contig of 25490 bp in length
* 51912
52011: gap of unknown length
* 52012
90309: contig of 38298 bp in length.
Location/Qualifiers
1. 90309
/organism="Takifugu rubripes"
/db_xref="taxon:31033"
/clone="205P11"
/clone_lib="Incyte Genomics"
1. 2330
/note="assembly_fragment"
2431..5020
/note="assembly_fragment"
5121..7685
/note="assembly_fragment"
7786..13195
/note="assembly_fragment"
13296..20253
/note="assembly_fragment"
20354..26321
/note="assembly_fragment"
26422..51911

FEATURES

source

misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature

/note="assembly_fragment"
clone_end:T7
vector_side:left
52012..90309
/note="assembly_fragment"
clone_end:SP6
vector_side:right
23869 a 20700 c 20314 g 24692 t 734 others
BASE COUNT
ORIGIN

Query Match 54.5%; Score 24; DB 2; Length 90309;
Best Local Similarity 75.0%; Pred. No. 14;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Oy 4 cttcaattgtgacaaagctgccaagagcgttgagtagga 43
Db 32119 CCTTCTGTATGGACCGCTGCCTACAGCTAAAGCAGGA 32158

RESULT 14

AC046194 AC046194 157241 bp DNA linear HTG 23-SEP-2000
LOCUS Homo sapiens chromosome 4 clone RP11-578A19 map 4, WORKING DRAFT
DEFINITION SEQUENCE, 13 unordered pieces.
AC046194
VERSION AC046194.3 GI:10280914
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE

1 (bases 1 to 157241)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 4, clone RP11-578A19
Unpublished
2 (bases 1 to 157241)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
Boguslavsky, L., Bouckhgalter, B., Brown, A., Burkett, G.,
Campiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
Gallagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Laroque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neil, D., Olivari, T.M., Oliver, J., Peterson, K., Plerre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, H., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigglio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE JOURNAL

COMMENT

Submitted (13-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 23, 2000 this sequence version replaced gi:8099902.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genom Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L9619
Center clone name: 578_A_19

Db 91972 TTTCCATTTTGGACGAGTTTTTTGATAGGCTTCAGTGGAG 92011
||||| ||| ||||| || | | ||||| ||||| |||||

Search completed: September 24, 2002, 15:53:42
Job time: 20317 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2002, 15:48:26 : Search time 403.26 Seconds
(without alignments)
187.334 Million cell updates/sec

Title: US-09-617-174B-1_COPY_914_957

Perfect score: 44

Sequence: 1 tctcttcaattgtgacaa.....ccaagaggcttgagtaggag 44

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*

- 1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44	100.0	1141	18	AAT89000 Human maspin promo
2	31.2	70.9	12426	22	AAS46481 Tumour suppressor
3	23	52.3	1820	16	AAQ88817 Human surfactant p
4	22.6	66.15	6615	23	ABL26203 Drosophila melanog
5	22.6	51.4	10565	23	ABL26202 Drosophila melanog
6	22.4	50.9	321	22	AAK50410 Human bone marrow
7	22.4	50.9	349	22	AAK54915 Human haematologic
8	22.4	50.9	527	22	AAK37539 Human bone marrow
9	22.4	50.9	12426	22	AAS46482 Tumour suppressor

C	10	22.2	50.5	2280	23	ABL26459	Drosophila melanog	
	11	22.2	50.5	4530	23	ABL26458	Drosophila melanog	
	12	22.2	50.5	8642	23	ABL02384	Drosophila melanog	
	13	22.2	50.5	74586	24	AAS16905	Genomic DNA encodi	
	14	22	50.0	1318	23	AAS71982	DNA encoding novel	
	C	15	21.6	49.1	500	22	ABA58422	Human foetal liver
	C	16	21.6	49.1	500	22	ABA27523	Probe #5989 for ge
	C	17	21.6	49.1	500	22	AAK06526	Human brain expres
	C	18	21.6	49.1	500	22	AAK32211	Human bone marrow
	C	19	21.6	49.1	500	22	AAI16008	Probe #5941 for ge
	C	20	21.6	49.1	500	22	AAI38062	Probe #6748 used t
	C	21	21.6	49.1	6585	21	AAA60446	Murine factor V en
	C	22	21.4	48.6	1422	22	AAH03825	Human reproductive
	C	23	21.4	48.6	1596	22	AAH34605	Human colon cancer
	C	24	21.4	48.6	1682	22	AAK71694	Human immune/haema
	C	25	21.4	48.6	1682	22	AAK71695	Human immune/haema
	C	26	21.4	48.6	1682	22	AAK71697	Human immune/haema
	C	27	21.4	48.6	2622	22	AAH18084	Human cDNA sequenc
	C	28	21	47.7	34279	22	AAF28539	Genomic fragment #
	C	29	21	47.7	140036	24	AAS98600	Human genomic DNA
	C	30	20.8	47.3	233	21	AAK21576	Human secreted pro
	C	31	20.8	47.3	458	22	AAK05085	Human brain expres
	C	32	20.8	47.3	458	22	AAK30641	Human bone marrow
	C	33	20.8	47.3	458	22	AAI36578	Probe #5264 used t
	C	34	20.8	47.3	966	24	ABI99550	Mouse ischaemic co
	C	35	20.8	47.3	2468	22	AAI59786	Human polynucleoti
	C	36	20.8	47.3	2868	22	AAH15964	Human cDNA sequenc
	C	37	20.8	47.3	4253	18	AAH6172	Human MLN 51 nucle
	C	38	20.8	47.3	4253	22	AAI58000	Human polynucleoti
	C	39	20.8	47.3	8396	21	AAZ59353	Human STP2 (phenol
	C	40	20.6	46.8	1038	23	ABL13843	Drosophila melanog
	C	41	20.6	46.8	1254	15	AAQ55782	Segment of human g
	C	42	20.6	46.8	1254	15	AAQ58718	Human gamma-1 chai
	C	43	20.6	46.8	2252	23	ABL14548	Drosophila melanog
	C	44	20.6	46.8	2705	23	ABL05757	Drosophila melanog
	45	20.6	46.8	2943	23	ABL22984	Drosophila melanog	

ALIGNMENTS

RESULT 1

AAAT89000

ID AAT89000 standard; cDNA; 1141 BP.

XX AAAT89000;

XX DT 22-APR-1998 (first entry)

XX Human maspin promoter and partial coding cDNA.

XX Maspin; serpin; mammary epithelial cell; human; promoter; malignant;

XX tumour cell; treatment; prostate cancer; breast cancer; metastasis; ss.

XX Homo sapiens.

XX key

XX Location/Qualifiers

FT promoter

FT /..1141..

FT /*tag= a

FT /note= "maspin promoter region"

FT misc_signal

FT 443..449

FT /*tag= b

FT /note= "Putative regulatory element AP2"

FT misc_signal

FT 451..457

FT /*tag= c

FT /note= "Putative regulatory element AP1"

FT misc_signal

FT 485..490

FT /*tag= d

FT /note= "Putative regulatory element Ets"

FT misc_signal

FT 684..691

FT /*tag= e

FT /note= "Putative regulatory element HRE"

FT misc_signal

FT 846..851

ALIGNMENTS

RESULT 1

AAT89000

ID AAT89000 standard; cDNA: 1141 BP.

XX

AC AAT89000;

XX

XX 22-APR-1998 (first entry)

DT Human maspin promoter and partial coding cDNA.

XX

DE Maspin; serpin; mammary epithelial cell; human; promoter; malignant;

XX tumour cell; treatment; prostate cancer; breast cancer; metastasis; ss.

OS Homo sapiens.

XX

PH Key Location/Qualifiers

FT promoter 1..1141.

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

Db 321 TCAACTGTGACCAAGGCTGCCAGAGGCTGGAG 290
 |||| |||| | ||||| ||||| |||

RESULT 8
 AAK37539
 ID AAK37539 standard; DNA: 527 BP.
 XX
 AC AAK37539;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human bone marrow expressed single exon probe SEQ ID NO: 12096.
 XX
 KW Human: bone marrow expressed exon; gene expression analysis; probe;
 KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
 XX
 OS Homo sapiens.
 OS
 PN WO200157276-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00668.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 XX WPI; 2001-488900/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human bone marrow -
 XX
 XX Example 4; SEQ ID NO: 12096; 658pp + Sequence Listing; English.
 XX
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
 CC the probes of the invention.
 XX
 XX Sequence 527 BP; 101 A; 112 C; 204 G; 110 T; 0 other;

Query Match 50.9%; Score 22.4; DB 22; Length 527;
 Best Local Similarity 81.2%; Pred. No. 14;
 Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 7 tcaattgtgacaagctgccagagcttgag 38
 |||| |||| | ||||| ||||| |||

Db 228 tcaactgtgaccaggtgccagaggtgag 259

RESULT 9
 AAS46482/c
 ID AAS46482 standard; DNA: 12426 BP.
 XX
 AC AAS46482;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Tumour suppressor gene derived chemically modified sequence #204.
 XX

KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
 KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
 KW cytosine methylation; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200168912-A2.
 XX
 PD 20-SEP-2001.
 XX
 PF 15-MAR-2001; 2001WO-EP02955.
 XX
 PR 15-MAR-2000; 2000DE-1013847.
 PR 06-APR-2000; 2000DE-1019058.
 PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 XX WPI; 2001-602752/68.
 XX
 PT Fragments of chemically modified genes associated with tumour suppressor
 PT genes and oncogenes, useful in designing primers and probes for
 PT analysing diseases associated with cytosine methylation state e.g.
 PT cancer
 XX
 PS Claim 1; SEQ ID No 204; 27pp; English.
 XX
 CC The invention relates to a nucleic acid comprising a sequence of 18
 CC bases, of a segment of chemically pretreated DNA (cp DNA) e.g. with
 CC bisulphite, of genes associated with tumour suppression and
 CC oncogenes having a sequence taken from 536 (actually 533 since
 CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
 CC (SS) and sequences complementary to (SS). The nucleic acid may be a
 CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
 CC form part of a set of probes for detecting the cytosine methylation state
 CC and/or single nucleotide polymorphisms and also to be used in an
 CC array for analysing diseases associated with CpG dinucleotides e.g.
 CC cancers and tumours. The probes can also be used in a method for
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis
 CC and/or therapy of existing diseases or the predisposition to specific
 CC diseases, by analysing cytosine methylations. The parameters may be
 CC compared to another set of genetic and/or epigenetic parameters, the
 CC differences serving as basis for diagnosis and/or prognosis events which
 CC are disadvantageous to patients. The present sequence is one of the
 CC 533 genomic sequences derived from tumour suppressor genes and
 CC oncogenes. Sequences with even numbered Seq ID numbers are the
 CC complementary sequence of the corresponding odd numbered sequence (e.g.
 CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence
 CC is missing).
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 12426 BP; 3791 A; 137 C; 2517 G; 5981 T; 0 other;

Query Match 50.9%; Score 22.4; DB 22; Length 12426;
 Best Local Similarity 72.5%; Pred. No. 30;
 Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 tctctttcaattgtggacaagctgccagaggttgagta 40
 | ||||| |||| | ||||| |||| | ||| | |||

Db 7593 TTTCTTTCAATTATTAACAACACTACCAAAAAAACTTAATA 7554

RESULT 10
 ABL26459/c
 ID ABL26459 standard; DNA: 2280 BP.

```

XX AC ABL26459;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 30850.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX OS pharmaceutical; gene; ds.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX PI WPI; 2001-656860/75.
XX PS New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX PS Claim 1; SEQ ID NO 30850; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins
XX CC (ABB57737-ABB72072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX PS Sequence 2280 BP; 587 A; 757 C; 451 G; 485 T; 0 other;
XX SQ
XX
XX Query Match 50.5%; Score 22.2; DB 23; Length 2280;
XX Best Local Similarity 69.8%; Pred. No. 24;
XX Matches 30; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
XX
XX Qy 1 tctcttcaattgtgacaagctgccaaaggcttgatagga 43
XX |||| |||| | ||| | |||| || | ||||
XX Db 1537 TCTCATTGATCGGGAGAGGATGCCAGATGGGTGGATAGGA 1495
XX
XX RESULT 11
XX ABL26458
XX ID ABL26458 standard; DNA; 4530 BP.
XX AC ABL26458;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 30847.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX OS pharmaceutical; gene; ds.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.

```

```

XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX PI WPI; 2001-656860/75.
XX PS New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX PS Claim 1; SEQ ID NO 30847; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins
XX CC (ABB57737-ABB72072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX PS Sequence 4530 BP; 1094 A; 946 C; 1239 G; 1251 T; 0 other;
XX SQ
XX
XX Query Match 50.5%; Score 22.2; DB 23; Length 4530;
XX Best Local Similarity 69.8%; Pred. No. 29;
XX Matches 30; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
XX
XX Qy 1 tctcttcaattgtgacaagctgccaaaggcttgatagga 43
XX |||| |||| | ||| | |||| || | ||||
XX Db 1744 tctcattcagatcgaggagagatgccaaagatgggtgatagga 1786
XX
XX RESULT 12
XX ABL02384
XX ID ABL02384 standard; cDNA; 8642 BP.
XX AC ABL02384;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 1634.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX OS pharmaceutical; gene; ss.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX PI WPI; 2001-656860/75.
XX DR

```



```

DR P-PSDB: ABB58281.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions
XX
XX Claim 1; SEQ ID NO 1634; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLI16176-ABLI30511), expressed DNA
CC sequences (ABLI01840-ABLI16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 8642 BP; 2272 A; 1990 C; 1984 G; 2396 T; 0 other;
SQ
Query Match 50.5%; Score 22.2; DB 23; Length 8642;
Best Local Similarity 69.8%; Pred. NO. 33;
Matches 30; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
QY 1 tctcttcgaattgtgacaaagctgcgaagagcgttgagtagga 43
Db 8457 tctcattcagatcgaggagagatgcgaagatgggtgtagga 8499
RESULT 13
AAS16905
ID AAS16905 standard; DNA; 74586 BP.
XX
AC AAS16905;
XX
XX 25-FEB-2002 (first entry)
XX
XX Genomic DNA encoding human transporter polypeptide.
XX
XX Human; transporter; gene therapy; nucleic acid hybridisation; ds;
KW ionotropic glutamate receptor subfamily; single nucleotide polymorphism;
KW chromosome 9.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH 1651..69740
FT CDS
FT /*tag= a
FT /product= "Human transporter protein"
FT 1651..2349
FT /*tag= b
FT /number= 1
FT 2350..29157
FT /*tag= c
FT /*tag= c
FT /number= 1
FT /cons_splice= (5'site:NO, 3'site:YES)
FT replace(3248,A)
FT /*tag= d
FT /standard_name= "Single nucleotide polymorphism"
FT replace(9928,A)
FT /*tag= e
FT /standard_name= "Single nucleotide polymorphism"
FT replace(11387,C)
FT /*tag= f
FT /standard_name= "Single nucleotide polymorphism"
FT replace(11578,T)
FT /*tag= g
FT /standard_name= "Single nucleotide polymorphism"
FT replace(11731,G)
FT /*tag= h
FT /number= 3
FT

```

FT	variation	replace(44478,C)	FT	variation	replace(71332,T)
FT		/*tag= ag	FT		/*tag= be
FT	intron	/standard_name= "Single nucleotide polymorphism"	FT	variation	/standard_name= "Single nucleotide polymorphism"
FT		45484..52584	FT		replace(72153,G)
FT		/*tag= ah	FT		/*tag= bf
FT		/number= 3	FT	variation	/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(51524,G)	FT		replace(72711,K)
FT		/*tag= ai	FT		/*tag= bg
FT	exon	/standard_name= "Single nucleotide polymorphism"	FT	variation	/standard_name= "Single nucleotide polymorphism"
FT		52585..52730	FT		replace(74434,C)
FT		/*tag= aj	FT		/*tag= bh
FT	intron	/number= 4	FT		/standard_name= "Single nucleotide polymorphism"
FT		52731..57554	PN	WO200181413-A2.	
FT		/*tag= ak	XX		
FT		/number= 4			
FT	variation	replace(54016,T)			
FT		/*tag= al			
FT	variation	/standard_name= "Single nucleotide polymorphism"			
FT		replace(54405,A)			
FT		/*tag= am			
FT	variation	/standard_name= "Single nucleotide polymorphism"			
FT		replace(55007,C)			
FT		/*tag= an			
FT	variation	/standard_name= "Single nucleotide polymorphism"			
FT		replace(55156,G)			
FT		/*tag= ao			
FT	exon	/standard_name= "Single nucleotide polymorphism"			
FT		57555..57670			
FT		/*tag= ap			
FT		/number= 5			
FT	intron	57671..59239			
FT		/*tag= aq			
FT		/number= 5			
FT	exon	59240..59391			
FT		/*tag= ar			
FT		/number= 6			
FT	intron	59392..63553			
FT		/*tag= as			
FT		/number= 6			
FT	exon	63554..63718			
FT		/*tag= at			
FT		/number= 7			
FT	intron	63719..65069			
FT		/*tag= au			
FT		/number= 7			
FT	variation	replace(64177,C)			
FT		/*tag= av			
FT		/standard_name= "Single nucleotide polymorphism"			
FT	exon	65070..65149			
FT		/*tag= aw			
FT		/number= 8			
FT	intron	65150..69400			
FT		/*tag= ax			
FT		/number= 8			
FT		/cons_splice= (5'site:NO, 3'site:NO)			
FT	variation	replace(66196,C)			
FT		/*tag= ay			
FT		/standard_name= "Single nucleotide polymorphism"			
FT	variation	replace(66780,G)			
FT		/*tag= az			
FT		/standard_name= "Single nucleotide polymorphism"			
FT	variation	replace(69176,T)			
FT		/*tag= ba			
FT		/standard_name= "Single nucleotide polymorphism"			
FT	exon	69401..69737			
FT		/*tag= bb			
FT		/number= 9			
FT	variation	replace(70027,R)			
FT		/*tag= bc			
FT		/standard_name= "Single nucleotide polymorphism"			
FT	variation	replace(70419,G)			
FT		/*tag= bd			
FT		/standard_name= "Single nucleotide polymorphism"			

Query Match

50.5%; Score 22.2; DB 24; Length 74586;

Best Local Similarity 69.8%; Pred. No. 54;

Matches 30; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy

1 tctcttcaattgtgcaagctgccaagagcttgcaagcttgtagtagga 43

11 11111 11111 11 11111 11111 11111 1

Db

4029 tctcttcatgttagattgtttgccaagagcttctaagtagca 4071

RESULT 14

AAS71982

ID AAS71982 standard; cDNA; 1318 BP.

XX

AC AAS71982;

XX

DT 13-FEB-2002 (first entry)

XX

DE DNA encoding novel human diagnostic protein #7786.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US08631.

XX

PR 31-MAR-2000; 2000US-0540217.

XX

PR 23-AUG-2000; 2000US-0649167.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Drmanac RT, Liu C, Tang YT;

XX

DR WPI; 2001-639362/73.

DR

DR P-PSDB; ABG07795.

XX

PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity -

XX

PS Claim 1; SEQ ID No 7786; 103pp; English.

XX

CC The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical

us-09-617-174b-1_copy_914_957.rng

Wed Sep 25 08:14:20 2002

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2002, 14:45:15 : Search time 3309.28 Seconds
(without alignments)
179.435 Million cell updates/sec

Title: US-09-617-174b-1_COPY_914_957

Perfect score: 44

Sequence: 1 tctcttcaattgtgacaa.....ccaagagccttgtagag 44

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estcnu:*
5: em_estcov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	24	54.5	314	12	BH083050 RPCT-24-2
2	23.6	53.6	385	12	A2467314
3	23.6	53.6	605	12	A24449029
4	23.6	53.6	676	12	A2462769
5	23.2	53.7	186	12	CNS03C23
6	23.2	53.7	497	9	AA539675
7	23	52.3	515	12	A2144680
8	22.8	51.8	248	9	BB544774
9	22.8	51.8	471	9	AV720858
10	22.8	51.8	477	9	AW399699
11	22.8	51.8	492	12	A2077326
12	22.8	51.8	507	10	BG137080
13	22.8	51.8	522	12	A20764541
14	22.8	51.8	545	12	A2748037
15	22.8	51.8	647	12	BH358813
16	22.8	51.8	658	9	BB615382
17	22.8	51.8	686	12	BH110962

c 18	22.8	51.8	750	10	BG16583	BG16583	602677839
c 19	22.8	51.8	971	9	AL523317	AL523317	
c 20	22.8	51.8	2396	11	BC022504	BC022504	Hom sapi
c 21	22.6	51.4	380	10	BF196527	BF196527	248303 MA
c 22	22.6	51.4	513	9	A1340928	A1340928	q082405. x
c 23	22.6	51.4	804	10	BG873257	BG873257	602794102
c 24	22.4	50.9	831	12	B73112	B73112	RPCT11-11E2
c 25	22.4	50.9	826	10	BF965281	BF965281	602268196
c 26	22.4	50.9	840	12	A2207828	A2207828	SP-0135-A
c 27	22.4	50.9	856	10	B1262248	B1262248	602954030
c 28	22.4	50.9	1149	12	CNS02K00	AL201993	Tetradon
c 29	22.2	50.5	351	12	A0777266	A0777266	HS-2261_A
c 30	22.2	50.5	405	12	A0634417	A0634417	RPCT-11-4
c 31	22.2	50.5	533	12	A0482843	A0482843	RPCT-11-2
c 32	22.2	50.5	557	10	BC460704	BC460704	RS143326
c 33	22.2	50.5	585	10	BM008659	BM008659	603618528
c 34	22.2	50.5	620	12	A2020356	A2020356	RPCT-23-2
c 35	22.2	50.5	657	10	B1915756	B1915756	603184267
c 36	22.2	50.5	699	10	B1194845	B1194845	602943980
c 37	22.2	50.5	701	10	BF534836	BF534836	602048373
c 38	22.2	50.5	719	10	B1764717	B1764717	603051054
c 39	22.2	50.5	727	10	BE799125	BE799125	601588844
c 40	22.2	50.5	762	12	BH187319	BH187319	033_K_14-
c 41	22.2	50.5	762	12	CNS07RLJ	AL624265	T7 end of
c 42	22.2	50.5	801	10	BE736747	BE736747	601306024
c 43	22.2	50.5	803	10	B1823154	B1823154	603039596
c 44	22.2	50.5	806	10	B1838085	B1838085	603086668
c 45	22.2	50.5	815	10	BF536852	BF536852	602048440

ALIGNMENTS

RESULT 1	BH083050	314 bp	DNA	linear	GSS 18-JUL-2001
LOCUS	RPCT-24-236D16.TJ	RPCT-24	Mus musculus	genomic clone	RPCT-24-236D16
DEFINITION	DNA sequence.				
ACCESSION	BH083050	GI:14902647			
VERSION	BH083050.1	GI:14902647			
KEYWORDS	GSS.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akintrel,B., Levins,M., Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E., Russell,D., de Jong,P. and Fraser,C.M. Mouse BAC End Sequences from Library RPCT-24 Unpublished (1999)				
AUTHORS	Contact: Shaying Zhao				
TITLE	Department of Eukaryotic Genomics				
JOURNAL	The Institute for Genomic Research				
COMMENT	9712 Medical Center Dr., Rockville, MD 20850, USA				
	Tel: 301 838 0200				
	Fax: 301 838 0208				
	Email: szhao@tigr.org				
	Clones are derived from the mouse BAC library RPCT-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.tigr.org/tdb/bac/orderingframe.htm). BAC end plate: 236 row: D column: 16				
	Seq primer: SP6				
	Class: BAC ends.				
FEATURES	Location/Qualifiers				
source	1..314				
	/organism="Mus musculus"				
	/strain="C57BL/6J"				
	/db_xref="taxon:10090"				
	/clone="RPCT-24-236D16"				
	/clone_lib="RPCT-24"				

```

/sex="Male"
/cell_type="Spleen/Brain"
/Note="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI;
RCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."
BASE COUNT      71 a      34 c      62 g      147 t
ORIGIN

Query Match      54.5%; Score 24; DB 12; Length 314;
Best Local Similarity 75.0%; Pred. No. 47;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 5 ttcaattgtgacaagctgcagaagagcttgagtagag 44
    |||||  |||||  |||||  |||||  |||||  |||||
Db 60 TTTCATTTTGGACAGGTATCATGAGCTGTGACAGAG 99

RESULT 2
LOCUS      A2467314      385 bp      DNA      linear      GSS 04-OCT-2000
DEFINITION clone UUGC1M0278A11 R, DNA sequence.
ACCESSION  A2467314
VERSION     A2467314.1 GI:10625439
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 385)
AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Rellily
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D., Weiss,R.
TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
JOURNAL   Unpublished (2000)
COMMENT   Contact: Robert B. Weiss
            University of Utah Genome Center
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000 Std Error: 0.00
            Plate: 0278 row: A column: 11
            Seq primer: CACACAGAAACAGCATATGACC
            Class: plasmid ends
            High quality sequence stop: 385.
FEATURES
    source
        1..385
            location/Qualifiers
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="UUGC1M0278A11"
            /clone_1lb="Mouse 10kb plasmid UUGC1M library"
            /sex="Male"
            /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
            /note="Vector: pMDA2ny; Purified genomic DNA from M.
            musculus C57BL/6J (male) was obtained from the Jackson
            Laboratory Mouse DNA Resource
            (http://www.jax.org/resources/documents/dnares/). The DNA
            was hydridynamically sheared by repeated passage through a
            0.005 inch orifice at constant velocity. The sheared DNA
            was blunt end-repaired with T4 DNA polymerase and T4
            polynucleotide kinase. Adaptor oligonucleotides were
            ligated to the blunt ends in high molar excess. The
            adaptor DNA was purified and size-selected for a 9.5 to
            10.5 kb range using preparative agarose gel

```

```

electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (g114732114/9b/AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor to mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT      100 a      94 c      58 g      133 t
ORIGIN

Query Match      53.6%; Score 23.6; DB 12; Length 385;
Best Local Similarity 76.3%; Pred. No. 71;
Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 6 ttcaattgtgacaagctgcagaagagcttgagtaga 43
    |||||  |||||  |||||  |||||  |||||  |||||
Db 191 TTTCATCTGTATCACTGCCAATATGCTTGATAGGA 228

RESULT 3
LOCUS      A2449029      605 bp      DNA      linear      GSS 04-OCT-2000
DEFINITION clone UUGC1M0247N06 F, DNA sequence.
ACCESSION  A2449029
VERSION     A2449029.1 GI:10602411
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 605)
AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Rellily
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D., Weiss,R.
TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
JOURNAL   Unpublished (2000)
COMMENT   Contact: Robert B. Weiss
            University of Utah Genome Center
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000 Std Error: 0.00
            Plate: 0247 row: N column: 06
            Seq primer: CGTTGTAAACGACGGCCACT
            Class: plasmid ends
            High quality sequence stop: 605.
FEATURES
    source
        1..605
            location/Qualifiers
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="UUGC1M0247N06"
            /clone_1lb="Mouse 10kb plasmid UUGC1M library"
            /sex="Male"
            /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
            /note="Vector: pMDA2ny; Purified genomic DNA from M.
            musculus C57BL/6J (male) was obtained from the Jackson
            Laboratory Mouse DNA Resource
            (http://www.jax.org/resources/documents/dnares/). The DNA
            was hydridynamically sheared by repeated passage through a
            0.005 inch orifice at constant velocity. The sheared DNA
            was blunt end-repaired with T4 DNA polymerase and T4
            polynucleotide kinase. Adaptor oligonucleotides were
            ligated to the blunt ends in high molar excess. The
            adaptor DNA was purified and size-selected for a 9.5 to
            10.5 kb range using preparative agarose gel

```

10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (q11473211419b1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 246 a 140 c 87 g 132 t
ORIGIN

Query Match 53.6% Score 23.6; DB 12; Length 605;
Best Local Similarity 76.3% Pred. No. 80;
Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

OY 7 tcaattgtgacaagctgccaaagagctgtagtagag 44
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 605 TCAATTTTGAGAGAGCTACCATGAGCTGTGAGAGAG 568

RESULT 4
A2462769/c 676 bp DNA linear GSS 04-OCT-2000
LOCUS 1M0271E17F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC1M0271E17 F, DNA sequence.
ACCESSION A2462769
VERSION A2462769.1 GI:10620894
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 676)
AUTHORS Dun, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A. and Wright, D., Weiss, R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts

JOURNAL Unpublished (2000)

COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert length: 1000 Std Error: 0.00
Plate: 0271 row: E column: 17
Seq primer: CTTGTAAACGACGCGCACT
Class: plasmid ends
High quality sequence stop: 676.
Location/Qualifiers

FEATURES
source 1..676
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0271E17"
/clone_11b="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pMD42m; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The

adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (q11473211419b1AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 283 a 159 c 89 g 145 t
ORIGIN

Query Match 53.6% Score 23.6; DB 12; Length 676;
Best Local Similarity 76.3% Pred. No. 83;
Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

OY 7 tcaattgtgacaagctgccaaagagctgtagtagag 44
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 650 TCAATTTTGAGAGAGCTACCATGAGCTGTGAGAGAG 613

RESULT 5
CNS03C23/c 186 bp DNA linear GSS 15-MAY-2000
LOCUS Tetradon nigroviridis genome survey sequence T7 end of clone
DEFINITION 015F06 of library G from Tetradon nigroviridis, genomic survey sequence.
ACCESSION AL238440
VERSION AL238440.1 GI:7897575
KEYWORDS GSS: genome survey sequence.
SOURCE Tetradon nigroviridis.
ORGANISM Tetradon nigroviridis

REFERENCE 1 (bases 1 to 186)
AUTHORS Roest-Crollius, H., Jallion, O., Dasilva, C., Fitzames, C., Fisher, C., Bouneau, L., Billault, A., Quetler, F., Saulin, W., Bernot, A. and Weissenbach, J.

TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetradon nigroviridis

JOURNAL Unpublished

COMMENT 2 (bases 1 to 186)
Roest-Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fitzames, C., Wincker, P., Brotlier, P., Quetler, F., Saulin, W. and Weissenbach, J.

TITLE Human gene number estimate provided by genome wide analysis using Tetradon nigroviridis DNA sequence

JOURNAL Unpublished
REFERENCE 3 (bases 1 to 186)
Genoscope.
Direct Submission
Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases
This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetradon nigroviridis genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetradon.
Location/Qualifiers

FEATURES
source 1..186
/organism="Tetradon nigroviridis"
/db_xref="taxon:99883"
/clone="015F06"
/clone_11b="G"
/note="Genoscope sequence ID : CNS015DC03LP1-end : T7"

BASE COUNT 59 a 40 c 54 g 31 t 2 others
ORIGIN

Query Match 53.2% Score 23.4; DB 12; Length 186;
Best Local Similarity 73.2% Pred. No. 69;

RESULT	9
AV720856/c	
LOCUS	AJ720858 471 bp mRNA linear EST 16-OCT-2000
DEFINITION	AJ720858 GLC Homo sapiens cDNA clone GICDC07 5', mRNA sequence.
ACCESSION	AJ720858
VERSION	AJ720858.1 GI:10818010
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
Eukaryota; Metazoa;	Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria;	Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 471)

```
FEATURES
SOURCE
location/Qualifiers
1..477
/organism="lycopersicon pennellii"
/db_xref="taxon:28526"
/clone="CLP18B12"
/clone_lib="L. pennellii trichome, Cornell University"
/tissue_type="trichome"
```

```

/dev stage="mixed stages"
/lab host="SOLR"
/note="Vector: pBluescript SK-, site 1: EcoRI; site 2:
XhoI; leaves of various stages were shaken in liquid
nitrogen, shearing off trichomes. This procedure yielded a
mixture of cells highly enriched for trichomes, with minor
contamination by other types of leaf cells."
BASE COUNT      152 a      92 c      102 g      131 t
ORIGIN

```

Query Match	Similarity	Score	DB	Length
Best Local	71.4%	Pred. No. 1,5e+02	93	477
Matches	30; Conservative	0; Mismatches	12; Indels	0; Gaps

QY	3	tcttccaatgtgacacagctgcacagagccttgtagtagag	44
DB	93	tttttttCAATTGGGGGAATAAACCTTGAAGCAATGAGTACAG	134

RESULT	11				
LOCUS	AZ077326/C				
DEFINITION	AZ077326	492 bp	DNA	linear	GSS 31-MAR-2000
ACCESSION	RPCT-23-399N3.TJ	RPCT-23	Mus musculus	genomic clone	RPCT-23-399N3,
VERSION	AZ077326				
KEYWORDS	AZ077326.1	GI:7370225			
SOURCE	GSS.				
ORGANISM	house mouse.				
REFERENCE	Mus musculus.				
AUTHORS	Eukariyota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 492)				
TITLE	Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Aktiret				
JOURNAL	, B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Kroll,M., de Jong,P				
COMMENT	and Fraser,C.M.				
	Mouse PAC and Sequences from Library	RPCT-23			
	Unpublished (1995)				
	Other_GSSs:	RPCT-23-399N3.TV			

Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPECI-23. For BAC library availability, please contact Pliet de Jong (pliet@adelong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (inforesgen.com). BAC end page: http://www.tigr.org/tdb/Bac_ends/mouse/bac_end_intro.html
 Plate: 399 row: N column: 3
 Seg primer: SP6
 Class: BAC ends.

FEATURES	Location/Qualifiers
source	1..492

FEATURES	source	Location/Qualifiers
		1. .492
		/organism="Mus musculus"
		/strain="C57BL/6J"
		/db_xref="taxon:10090"
		/clone="RPC1-23-39M3"
		/clone_1lb="RPC1-23"
		/sex="Female"
		/lab_host="DH10B"
		/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site-1: ECORI; Site-2: ECOR1; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of Ecor1 and Ecor1 methylase. Size selected DNA was cloned into the pBACE3.6 vector at the ECORI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT	142 a	107 c 84 g 159 t
ORIGIN		

Query Match	51.8%	Score	22.8	DB	12	Length	493
Best local similarity	71.4%	Pred.	No. 1.5e+02				
Matches	30	Conservative	0	Mismatches	12	Indels	0
							Gaps 0
QY	2	ctcttcaattgtgtgacaaagctgcgaagaagctttagataga	43				
DB	238	ctcaactcctgtgtgacacatgttagcccaacagctttagacatga	197				

[illegible]

REFERENCE AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 507) van der Hoeven, R., Bezzerides, J., Sun, H., Cho, J., Uterback, T., Hansen, C., Rönning, C. and Tanksley, S. Generation of ESTs from wild tomato (<i>L. pennellii</i>) pollen Unpublished (2001) Contact: CUGI Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Email: http://www.genome.clemson.edu/orders/index.html .			

FEATURES					
source					
	1..507	Location/Qualifiers			
	/organism="Lycopersicon pennellii"				
	/cultivar="TAS6"				
	/db_xref="taxon:28526"				
	/clone="CLPP4P19"				
	/clone_lib="wild tomato pollen"				
	/tissue_type="pollen"				
	/dev_stage="pollen collected from open flowers"				
	/lab_host="SOLR"				
	/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:				
	XhoI; Pollen was collected from open flowers from				
	L.pennellii TAS6, and stored at -80 C until library				
	construction."				
BASE COUNT	164 a	95 c	122 g	126 t	
ORIGIN					
Query Match	51.8%;	Score 22.8;	DB 10;	Length 507;	
Best Local Similarity	71.4%;	Pred. No. 1.6e+02;			
Matches	30;	Conservative	0;	Mismatches 12;	Indels 0;
				Gaps	0
DQ	3	tctttcaattgtgacaagctgcacaaaggcttagtagagag	44		
Dd	37	TTTTTCATTGGGGGAAATTAACCTAGAACAGCATGTAGACAG	78		

RESULT	13		
LOCUS	A0764541	522 bp	DNA
DEFINITION	HS 3168. A1 H04.MR CIT Approved Human Genomic Sperm Library D Homo		
ACCESSION	A0764541	satiens genomic clone	plate=3168 col=7 Row=O, DNA sequence.
VERSION	A0764541.1	GI:5642657	
KEYWORDS	GSS.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		


```
/clone="CH230-163024"  
/clone.lib="CHORI-230 Segment 1"  
/sex="Female"  
/cell_type="Brain"  
/note="Vector: PTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;  
CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by  
Pieter de Jong"  
BASE COUNT      211 a      119 c      132 g      185 t  
ORIGIN
```

```
Query Match      51.8%; Score 22.8; DB 12; Length 647;  
Best Local Similarity 79.4%; Pred. No. 1.7e+02;  
Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
QY 1 tctcttcaattgtgacaagctgccaagagct 34  
    ||| ||||| || ||| ||||| ||||| |||  
Db 488 TCTATTTCATTTTTGGTCAAGCAGCCAAAGGAT 455
```

Search completed: September 24, 2002, 14:45:19
Job time: 16654 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: September 24, 2002, 13:52:20 ; Search time 1872.72 Seconds
(without alignments)
12749.999 Million cell updates/sec

Title: US-09-617-174B-1
Perfect score: 1141
Sequence: 1 agataagcacagcagagaag.....agdataactgtgactccagg 1141

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vl.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vl.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
------------	-------	-------	-------	--------	----	-------------

1	805	70.6	170269	9	AC090307	AC090307 Homo sapi
2	805	70.6	173131	2	AC036176	AC036176 Homo sapi
3	805	70.6	173836	2	AP001404	AP001404 Homo sapi
4	337	29.5	214212	2	AC015793	AC015793 Homo sapi
5	226	19.8	214212	2	AC015793	AC015793 Homo sapi
6	89	7.8	1289	9	BC020713	BC020713 Homo sapi
7	61	5.3	2566	6	AX34331	AX34331 Sequence
8	61	5.3	2566	9	HSU04313	U04313 Human maspi
9	61	5.3	2584	6	AR036974	AR036974 Sequence
10	61	5.3	2584	6	AR070462	AR070462 Sequence
11	61	5.3	2584	6	115826	115826 Sequence 1
12	61	5.3	2584	6	174315	174315 Sequence 13
13	60	5.3	164512	9	AL365181	AL365181 Human DNA
14	57	5.0	179307	2	AC013787	AC013787 Homo sapi
15	57	5.0	208777	2	AC013391	AC013391 Homo sapi
16	55	4.8	158427	9	AC002553	AC002553 Homo sapi
17	53	4.6	2137	6	AX207159	AX207159 Sequence
18	52	4.6	140416	2	AC093024	AC093024 Homo sapi
19	52	4.6	156795	2	AC022290	AC022290 Homo sapi
20	52	4.6	172272	2	AC090671	AC090671 Homo sapi
21	52	4.6	182084	2	AC026615	AC026615 Homo sapi
22	50	4.4	81410	9	AP004715	AP004715 Homo sapi
23	50	4.4	148555	2	AP004711	AP004711 Homo sapi
24	50	4.4	160887	2	AC021498	AC021498 Homo sapi
25	50	4.4	207739	2	AC012564	AC012564 Homo sapi
26	49	4.3	12426	6	AX251236	AX251236 Sequence
27	48	4.2	31720	9	HSL196E3	268277 Human DNA s
28	48	4.2	172800	2	AL356597	AL356597 Homo sapi
29	48	4.2	183430	9	AL391827	AL391827 Human DNA
30	48	4.2	196476	2	AC055890	AC055890 Homo sapi
31	48	4.2	199255	2	AC084195	AC084195 Homo sapi
32	47	4.1	151791	2	AC093532	AC093532 Homo sapi
33	47	4.1	159906	2	AL590706	AL590706 Homo sapi
34	47	4.1	172004	9	AC008482	AC008482 Homo sapi
35	47	4.1	196606	9	AP000779	AP000779 Homo sapi
36	47	4.1	211967	2	AC013368	AC013368 Homo sapi
37	47	4.1	225157	2	AC027281	AC027281 Homo sapi
38	47	4.1	240579	2	AL360154	AL360154 Homo sapi
39	47	4.1	245202	2	AC092369	AC092369 Homo sapi
40	46	4.0	165560	9	AC068725	AC068725 Homo sapi
41	46	4.0	231972	2	AC068055	AC068055 Homo sapi
42	45	3.9	101611	9	AC018706	AC018706 Homo sapi
43	45	3.9	107017	9	AC012510	AC012510 Homo sapi
44	45	3.9	154306	2	AC034114	AC034114 Homo sapi
45	45	3.9	158905	9	AC005666	AC005666 Homo sapi

ALIGNMENTS

RESULT	1	AC090307	170269 bp	DNA	linear	PRI 30-DEC-2001
LOCUS	AC090307	Homo sapiens chromosome 18, clone RP11-851B10, complete sequence.				
DEFINITION	AC090307	HTG.				
ACCESSION	AC090307.7	GI:16041402				
VERSION	HTG.					
KEYWORDS	human.					
SOURCE	Homo sapiens					
ORGANISM	Homo sapiens					
REFERENCE	1	(bases 1 to 170269)				
AUTHORS	Birren, B., Linton, L., Nusbaum, C. and Lander, E.					
TITLE	Homo sapiens chromosome 18, clone RP11-851B10					
JOURNAL	Unpublished					
REFERENCE	2	(bases 1 to 170269)				
AUTHORS	Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S., Bana, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginç, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,					

Wed Sep 25 08:14:15 2002

Jones, C., Karatas, A., LaRocque, K., Lamaze, R., Landers, T.,
Lelenczky, J., Levine, R., Liu, G., MacLean, C., Macdonald, P.,
Mehoczky, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K.,
McPheters, R., Meldrim, J., Menius, L., Mihova, T., Mlenga, V.,
Munro, P., Nguyen, C., Norbu, C., Norman, C.H.,
Taylor, J., Nguyen, C., O'Neill, D., Oliver, J., Peterson, K.,
O'Donnell, P., O'Donnell, P., Pollara, V., Raymond, C., Retta, R.,
Phunkhang, P., Pierre, N., Pollara, V., Roman, J., Rosetti, M.,
Rieback, M., Riley, R., Rise, C., Roqov, P., Roman, J., Rosetti, M.,
Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P.,
Soungez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Travers, M., Travis, N., Triggillo, J., Vassiliou, H., Viel, R., VO, A.,
Wilson, B., Wu, X., Wyman, D., Ye, W.-J., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A., and Zody, M.

TITLE Direct Submission
JOURNAL Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 170269)

3 (bases 1 to 170269)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campoliano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collingmore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferrelira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Hearford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A., Karatas, A., Kells, C., LaRoque, K., Lamasares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Meng, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zia, R., Zembek, L., Zimmer, A., and Zody, M.

TITLE
JOURNAL
COMMENT

On Oct 11, 2001 this sequence was identified using RepeatMasker:
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research

Center: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information
Center project name: L12576
Center clone name: 851 B 10

Center Name	Location/Qualifiers
1. 170369	

```
/organism="Homo sapiens"  
/db_xref="taxon:9606"  
/chromosome="18"
```

```
/map="18"  
/clone="Rp11-851B10"  
/clone_lib="RPCI-11 Human Male BAC"
```

```

region complement(2. .177)
      /rpt_family="L1MC5"
region complement(189. .489)

```

```

    /rpt_family="AluSx"
    complement(708..803)
    /rpt_family="MIR"

```

```
complement(1616. .1904)
/rpt_family="AluJo"
complement(2815. .3120)
```

repeat_region	/rpt_family="AluSx"	3218..3365
repeat_region	/rpt_family="MIR"	3565..3590
repeat_region	/rpt_family="(TG)n"	complement(3782..4028)
repeat_region	/rpt_family="MIR"	4056..4117
repeat_region	/rpt_family="MERSA"	4601..5620
repeat_region	/rpt_family="HSMAR2"	5621..5643
repeat_region	/rpt_family="(CAA)n"	5644..5866
repeat_region	/rpt_family="HSMAR2"	6184..6475
repeat_region	/rpt_family="L1MA2"	6677..6837
repeat_region	/rpt_family="(TA)n"	6852..7018
repeat_region	/rpt_family="(TA)n"	6994..7122
repeat_region	/rpt_family="(CATATA)n"	7146..7332
repeat_region	/rpt_family="(CATATA)n"	7371..7526
repeat_region	/rpt_family="(TA)n"	complement(7527..7835)
repeat_region	/rpt_family="AluSx"	7936..8080
repeat_region	/rpt_family="MIR"	8697..8927
repeat_region	/rpt_family="AluSx"	complement(9036..9288)
repeat_region	/rpt_family="AluSq"	9873..9934
repeat_region	/rpt_family="(GA)n"	10340..10367
repeat_region	/rpt_family="AT-rich"	10368..10686
repeat_region	/rpt_family="AluY"	10687..10711
repeat_region	/rpt_family="AT-rich"	complement(10870..11043)
repeat_region	/rpt_family="L2"	11348..11375
repeat_region	/rpt_family="(T)n"	12359..12449
repeat_region	/rpt_family="L3"	complement(12634..12890)
repeat_region	/rpt_family="MIR"	complement(13436..13725)
repeat_region	/rpt_family="AluJb"	14244..14297
repeat_region	/rpt_family="(CGGG)n"	complement(15186..15350)
repeat_region	/rpt_family="MIR3"	16056..16114
repeat_region	/rpt_family="MIR3"	complement(16129..16429)
repeat_region	/rpt_family="AluSx"	16532..16646
repeat_region	/rpt_family="MER94"	16855..17145
repeat_region	/rpt_family="L2"	17377..17572
repeat_region	/rpt_family="MIR3"	complement(20036..20330)
repeat_region	/rpt_family="Charlie1"	complement(20333..20486)
repeat_region	/rpt_family="Charlie1"	20625..20691
repeat_region	/rpt_family="L2"	

Qy	622	gggcagctacatctctcttggcaactagacagaatagcgtgctgcagtttacacaaaagaat	681
Db	13897	GGCAGCTACCTCTCTTTGGCATCTAGCAGAAATGAGCTGCTCAGATTTACACAAAAGAAT	13956
Qy	682	ggaatcagactacttttggccaccaacgctctcagaagaattgtagttactatca	741
Db	13957	GGAGATCAGAGTACTTTTGTGGCCACCACCGTGTGAGAAATTTGTAGTGTACTATCA	14016
Qy	742	tcacacattactttatttcacgaatatttcaccttcgcgtcctgctggtggccgagagg	801
Db	14017	TCACACATTACTTTATTTTCATCGAATATTTTCACCTTCCGGTCTCGGTGGCCGAGAGG	14076
Qy	802	attgcgtacgcatgtctgtcatgtatgcatgaaactcacagccccctctctgcccgaaat	861
Db	14077	ATTGCCGTACGCATGCTGTGACGTATGATGTAACCTACAGCCCTTCTCTGCCGGAACAT	14136
Qy	862	gttagaggccttttggaaactgtgcagacaacagcaacttcagctgaatcatctcttc	921
Db	14137	GTTGGAGGCCCTTTTGGAAAGCTGTGCAGACAACAGTAACCTCAGCGTGAATCATTTCTTTC	14196
Qy	922	aattgtggacaagctgccaaagagctgttagtagagagagagtgctgcgcgagagcggggcgg	981
Db	14197	ATTGTGTCGAACAGTGCCCAAGAGCTTGAGTAGAGAGGAGTGGCCGCGAGCGGGGGCGG	14256
Qy	982	ggcggggcgctggagctgggctggcagtgggcgctggcgctggcgctgccagggtgagccaccgc	1041
Db	14257	GGCGGGGCGTGGAGCTGGGCTGGCAGTGGCGCTGGCGGTGCTGCCAGTGCAGCACCGC	14316
Qy	1042	tgccttcgccagacacaggtgcgctccacatccaggctcttggctcctcgtctgcctgt	1101
Db	14317	TGCTTCTGCCCAGACACGGTCCGCTCCACATCCAGTCTTTTGTGCTCCTCGCTTGGCTGT	14376
Qy	1102	tcctttccacgcattttccaggataactgtgactccagg	1141
Db	14377	TCCTTTTCCACGCATTTTCCAGGATACTGTGACTCCAGG	14416
RESULT 2			
AC036176/c			
LOCUS			
DEFINITION			
AC036176			
ACCESSION			
VERSION			
AC036176.7 GI:17647019			
KEYWORDS			
HTG; HTGS_PHASE3; HTGS_FULLTOP; HTGS_ACTIVEFIN.			
SOURCE			
ORGANISM			
human.			
Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi			
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
LOCUS			
AC036176			
DEFINITION			
Homo sapiens chromosome 18 clone NP11-635N19 map 18, *** SEQUENCE			
IN PROGRESS ***, 2 ordered pieces.			
ACCESSION			
AC036176			
VERSION			
AC036176.7 GI:17647019			
KEYWORDS			
HTG; HTGS_PHASE3; HTGS_FULLTOP; HTGS_ACTIVEFIN.			
SOURCE			
human.			
Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi			
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			

COMMENT

On Jul 14, 2000 this sequence version replaced gl:8117315.
----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsr.riken.go.jp
----- Project Information
Center project name: HumDraft18
Center clone name: RP11-851B10
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator Er-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 171265 bases at least Q40
Consensus quality: 172100 bases at least Q30
Consensus quality: 172469 bases at least Q20
Insert size: 172836; sum-of-contigs
Quality coverage: 10.50x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of
11 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved

1 68896 contig of 68896 bp in length
68997 95899 contig of 26903 bp in length
96000 119268 contig of 23269 bp in length
119369 141739 contig of 22371 bp in length
141840 152655 contig of 10816 bp in length
152756 159466 contig of 6711 bp in length
159567 163465 contig of 3899 bp in length
163566 166433 contig of 2868 bp in length
166534 170112 contig of 3579 bp in length
170213 172095 contig of 1883 bp in length
172196 173836 contig of 1641 bp in length.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 68896: contig of 68896 bp in length
* 68897 68996: gap of 100 bp
* 68997 95899: contig of 26903 bp in length
* 95900 95999: gap of 100 bp
* 96000 119268: contig of 23269 bp in length
* 119269 119368: gap of 100 bp
* 119369 141739: contig of 22371 bp in length
* 141740 141839: gap of 100 bp
* 141840 152655: contig of 10816 bp in length
* 152656 152755: gap of 100 bp
* 152756 159466: contig of 6711 bp in length
* 159467 159566: gap of 100 bp
* 159567 163465: contig of 3899 bp in length
* 163466 163565: gap of 100 bp
* 163566 166433: contig of 2868 bp in length
* 166434 166533: gap of 100 bp
* 166534 170112: contig of 3579 bp in length
* 170113 170212: gap of 100 bp
* 170213 172095: contig of 1883 bp in length
* 172096 172195: gap of 100 bp
* 172196 173836: contig of 1641 bp in length.
* Location/Qualifiers
1. 173836
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18q22"
/clone="RP11-851B10"

FEATURES
source

misc_feature 1. 68896
/note="assembly_fragment"
misc_feature 68997..95899
/note="assembly_fragment clone_end:T7 vector_side:right"
misc_feature 96000..119268
/note="assembly_fragment"
misc_feature 119369..141739
/note="assembly_fragment"
misc_feature 141840..152655
/note="assembly_fragment"
misc_feature 152756..159466
/note="assembly_fragment"
misc_feature 159567..163465
/note="assembly_fragment"
misc_feature 163566..166433
/note="assembly_fragment clone_end:SP6 vector_side:left"
misc_feature 166534..170112
/note="assembly_fragment"
misc_feature 170213..172095
/note="assembly_fragment"
misc_feature 172196..173836
/note="assembly_fragment"
BASE COUNT 53047 a 34410 c 33807 g 51570 t 1002 others
ORIGIN

Query Match 70.6%; Score 805; DB 2; Length 173836;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1055; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 82 ccaaatgaagaagctgtggaagacagagagacagagagctccacagagagatttca 141
|||||
Db 150643 CCAATGAAGAAGCTGTGGAGACAGGAGACAGAGCGCTCCACGAGAGATTTC 150584
Qy 142 gagcagagctgcgtactctcttttttttttttttttttttttttttttttttt 201
Db 150583 GAGCAGAGCTGCTACTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTT 150524
Qy 202 tacagtgttagctcacgctcacgctcacgctcacgctcacgctcacgctcacgct 261
|||||
Db 150523 TACAGTGTGTAGCTCACGCTCACGCTCACGCTCACGCTCACGCTCACGCTC 150464
Qy 262 gtctcagcttttccaagtaactggggaccacagcagctacacacagcagctaggtattgttt 321
Db 150463 GTCTCAGCTTTTCCAAGTAACTGGGACCACAGCGCATGTCATCCACACATAGGCTATTGTTT 150404
Qy 322 tacatttttttagagatgggggtctcacacattgtccacaggttggtctcaaaactcctggg 381
Db 150403 TACATTTTGTGAGAGTGGGGTCTCAGCATGTTGCCAGGTTGGTCTCAAACTCCTGGG 150344
Qy 382 ctcaagcaatccgctcacgctcaaacctcccaaatgctgggattacagcgctgagccacg 441
Db 150343 CTCAAGCAATCCGCTCACGCTCAACCTCCCAATGCTGGATTACGGGCTGAGCCACCG 150284
Qy 442 ggcagggctgagtaactcctaatacagcagatttttaaaagaacctcttcgcgccaccat 501
Db 150283 CGCCAGGCTGAGTAATCTTAATCAGAGATTTTAAAGAAACTTCTTGCCGCCACCAT 150224
Qy 502 taacaatactctccaccacatttgtagtaataattttgctaatactacacacacacacacacac 561
Db 150223 TAAACAATATCTCCCTACCAATTTGGTAGTAATAATTTTGTCTAATAGTACCTAATTTTGTAG 150164
Qy 562 gtggcactggtttattacatatatccattcctcttttttttttttttttttttttttttttttt 621
Db 150163 GTAGGCACTGTGTTTATACATATATCCATTCCTTTTGTATTGCTCTTTCTGTTTAAAT 150104
Qy 622 ggcagctacctctcttggtcattctagcagaatgagctgctgagcagtttacacacacacacacacac 681
Db 150103 GGGCAGCTACCTCTCTTGCCATCTAGCAGATGAGCTGCTGAGTTTACACAAAAGAAAT 150044
Qy 682 ggagatcagagctactctttttgtgccaccacacgctgctgagaaaatttgtagtgtaatacata 741
Db 150043 GGAGATCAGAGTACTTTTGTGTCACCAACGCTGCTGAGAAAATTGTAGTGTACTATCA 149984


```

* 23500 24217: contig of 718 bp in length
* 24218 24317: gap of 100 bp
* 24318 25026: contig of 709 bp in length
* 25027 25126: gap of 100 bp
* 25127 25836: contig of 710 bp in length
* 25837 25936: gap of 100 bp
* 25937 26637: contig of 701 bp in length
* 26638 26737: gap of 100 bp
* 26738 27446: contig of 709 bp in length
* 27447 27546: gap of 100 bp
* 27547 28274: contig of 728 bp in length
* 28275 28374: gap of 100 bp
* 28375 29095: contig of 721 bp in length
* 29096 29195: gap of 100 bp
* 29196 29917: contig of 722 bp in length
* 29918 30017: gap of 100 bp
* 30018 30724: contig of 707 bp in length
* 30725 30824: gap of 100 bp
* 30825 31529: contig of 705 bp in length
* 31530 31629: gap of 100 bp
* 31630 32324: contig of 695 bp in length
* 32325 32424: gap of 100 bp
* 32425 33116: contig of 692 bp in length
* 33117 33216: gap of 100 bp
* 33217 33936: contig of 720 bp in length
* 33937 34036: gap of 100 bp
* 34037 34750: contig of 714 bp in length
* 34751 34850: gap of 100 bp
* 34851 35578: contig of 728 bp in length
* 35579 35678: gap of 100 bp
* 35679 36393: contig of 715 bp in length
* 36394 36493: gap of 100 bp
* 36494 37203: contig of 710 bp in length
* 37204 37303: gap of 100 bp
* 37304 37965: contig of 662 bp in length
* 37966 38065: gap of 100 bp
* 38066 38781: contig of 716 bp in length
* 38782 38881: gap of 100 bp
* 38882 39590: contig of 709 bp in length
* 39591 39690: gap of 100 bp
* 39691 40403: contig of 713 bp in length
* 40404 40503: gap of 100 bp
* 40504 41238: contig of 735 bp in length
* 41239 41338: gap of 100 bp
* 41339 42064: contig of 726 bp in length
* 42065 42164: gap of 100 bp
* 42165 42880: contig of 716 bp in length
* 42881 42980: gap of 100 bp
* 42981 43687: contig of 707 bp in length
* 43688 43787: gap of 100 bp
* 43788 44504: contig of 717 bp in length
* 44505 44604: gap of 100 bp
* 44605 45307: contig of 703 bp in length
* 45308 45407: gap of 100 bp
* 45408 46106: contig of 699 bp in length
* 46107 46206: gap of 100 bp
* 46207 46909: contig of 703 bp in length
* 46910 47009: gap of 100 bp
* 47010 47696: contig of 687 bp in length
* 47697 47796: gap of 100 bp
* 47797 48531: contig of 735 bp in length
* 48532 48631: gap of 100 bp
* 48632 49341: contig of 710 bp in length
* 49342 49441: gap of 100 bp
* 49442 50162: contig of 721 bp in length
* 50163 50262: gap of 100 bp
* 50263 50990: contig of 728 bp in length
* 50991 51090: gap of 100 bp
* 51091 51803: contig of 713 bp in length
* 51804 51903: gap of 100 bp
* 51904 52610: contig of 707 bp in length
* 52611 52710: gap of 100 bp
* 52711 53412: contig of 702 bp in length

```

```

* 53413 53512: gap of 100 bp
* 53513 54253: contig of 741 bp in length
* 54254 54353: gap of 100 bp
* 54354 55084: contig of 731 bp in length
* 55085 55184: gap of 100 bp
* 55185 55972: contig of 788 bp in length
* 55973 56072: gap of 100 bp
* 56073 56855: contig of 783 bp in length
* 56856 56955: gap of 100 bp
* 56956 57708: contig of 753 bp in length
* 57709 57808: gap of 100 bp
* 57809 58579: contig of 771 bp in length
* 58580 58679: gap of 100 bp

Query Match      29.5%; Score 337; DB 2: Length 214212;
Best Local Similarity 99.3%; Pred. No. 1.2e-181;
Matches 537; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 82 ccaaatgaagaagctgtgggaagacagaggacagaagaacagcgtccacgaagagatttca 141
|||||
Db 168524 CCAAAATGAAGAAGCTGTGGGAAGACAGAGGAGACAAGACAGGCTCCACGAACAGATTTC 168465

QY 142 gagcagagctgcgtactccttttttttttttttttttttttttttttttttttttttt 201
|||||
Db 168464 GAGCAGAGCTGCGTACTCTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 168405

QY 202 tacagtgttagctcacgctcactgcagcttttgacctccacagcgtcaagtgtatctctc 261
|||||
Db 168404 TACAGTGGTTAGCTCACGCTCAGCTGACGCTTTTGACCTCCCAAGGCTCAAGTGATCCTCTC 168345

QY 262 gtctcagctttccaagtaactggggaccacagcagcatgcacacgctaggctattgttt 321
|||||
Db 168344 GTCTCAGCTTTCCAAAGTAACGTGGGACCACAGCAGCATGCATCACCACATAGGCTATTGTTT 168285

QY 322 tacatttttttagagatggggtctcaccatgttcccaggtgtgtctcaactcctcgtg 381
|||||
Db 168284 TACATTTTTTTGAGAGATGGGGTCTCACCATGTTGCCAGGTTGGTCTCAAACTCCTGGG 168225

QY 382 ctcaagcaatccgtcacgtcaacctccccaaatgctgggattacagcgtgagccaccg 441
|||||
Db 168224 CTCGAAGCAATCCGCTCAGCTCAACCTCCCAAAATGCTGGGATTACAGGCTGAGCCACCT 168165

QY 442 ggcagggtgagtaacttaataacagagatttttaaaagaacacttcctgcgccaccct 501
|||||
Db 168164 CGCCAGGCTGAGTAATCTTAATCAGAGATTTTAAAAAGAAACTTCTTCGCCACCACCAT 168105

QY 502 taacaataatctcaccacaaatttggtagtaataattttgctaaatagtaacctaatatttag 561
|||||
Db 168104 TAAACAATATCTCTTACCATAATTTGGTAGTAATAATTTTGTCTAATAGTACCTAATTTT 168045

QY 562 gtaggcactgttttatcatatataccattcctcttttttttttttttttttttttttttt 621
|||||
Db 168044 GTAGGCACGTGTTTATACATATATATCCATTCCTTCTTTTGTGATGCTCTTCTGTTTAA 167985

QY 622 g 622
Db 167984 G 167984

RESULT 5
AC015793
LOCUS
DEFINITION Homo sapiens chromosome 18 clone RP11-1117D15 map 18, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC015793
VERSION AC015793.2 GI:7144769
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (Bases 1 to 214212)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

```

TITLE JOURNAL REFERENCE AUTHORS

Homo sapiens chromosome 18, clone RP11-1117D15
Unpublished
2 (bases 1 to 214212)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Balwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhalter,B.,
Brown,A., Castle,A., Collange,M., Collins,S., Collymore,A.,
Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye-S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

TITLE JOURNAL COMMENT

Direct Submission
Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 3, 2000 this sequence version replaced gi:6446894.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence-submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1590
Center clone name: 1117_D15

NOTE: This record contains 240 individual
sequencing reads that have not been assembled into
contigs. Runs of N are used to separate the reads
and the order in which they appear is completely
arbitrary. Low-pass sequence sampling is useful for
identifying clones that may be gene-rich and allows
overlap relationships among clones to be deduced.
However, it should not be assumed that this clone
will be sequenced to completion. In the event that
the record is updated, the accession number will
be preserved.

1 695: contig of 695 bp in length
696 795: gap of 100 bp
796 1513: contig of 718 bp in length
1514 1613: gap of 100 bp
1614 2337: contig of 724 bp in length
2338 2437: gap of 100 bp
2438 3161: contig of 724 bp in length
3162 3261: gap of 100 bp
3262 3966: contig of 705 bp in length
3967 4066: gap of 100 bp
4067 4772: contig of 706 bp in length
4773 4872: gap of 100 bp
4873 5564: contig of 692 bp in length
5565 5664: gap of 100 bp
5665 6397: contig of 733 bp in length
6398 6497: gap of 100 bp
6498 7202: contig of 705 bp in length
7203 7302: gap of 100 bp
7303 8028: contig of 726 bp in length
8029 8128: gap of 100 bp
8129 8803: contig of 675 bp in length
8804 8903: gap of 100 bp
8904 9613: contig of 710 bp in length
9614 9713: gap of 100 bp
9714 10409: contig of 696 bp in length
10410 10509: gap of 100 bp
10510 11237: contig of 728 bp in length
11238 11337: gap of 100 bp
11338 12043: contig of 706 bp in length

12044 12143: gap of 100 bp
12144 12867: contig of 724 bp in length
12868 12967: gap of 100 bp
12968 13660: contig of 693 bp in length
13661 13760: gap of 100 bp
13761 14489: contig of 729 bp in length
14490 14589: gap of 100 bp
14590 15314: contig of 725 bp in length
15315 15414: gap of 100 bp
15415 16129: contig of 715 bp in length
16130 16229: gap of 100 bp
16230 16946: contig of 717 bp in length
16947 17046: gap of 100 bp
17047 17769: contig of 723 bp in length
17770 17869: gap of 100 bp
17870 18574: contig of 705 bp in length
18575 18674: gap of 100 bp
18675 19379: contig of 705 bp in length
19380 19479: gap of 100 bp
19480 20186: contig of 707 bp in length
20187 20286: gap of 100 bp
20287 20980: contig of 694 bp in length
20981 21080: gap of 100 bp
21081 21779: contig of 699 bp in length
21780 21879: gap of 100 bp
21880 22594: contig of 715 bp in length
22595 22694: gap of 100 bp
22695 23399: contig of 705 bp in length
23400 23499: gap of 100 bp
23500 24217: contig of 718 bp in length
24218 24317: gap of 100 bp
24318 25026: contig of 709 bp in length
25027 25126: gap of 100 bp
25127 25836: contig of 710 bp in length
25837 25936: gap of 100 bp
25937 26637: contig of 701 bp in length
26638 26737: gap of 100 bp
26738 27446: contig of 709 bp in length
27447 27546: gap of 100 bp
27547 28274: contig of 728 bp in length
28275 28374: gap of 100 bp
28375 29095: contig of 721 bp in length
29096 29195: gap of 100 bp
29196 29917: contig of 722 bp in length
29918 30017: gap of 100 bp
30018 30724: contig of 707 bp in length
30725 30824: gap of 100 bp
30825 31529: contig of 705 bp in length
31530 31629: gap of 100 bp
31630 32324: contig of 695 bp in length
32325 32424: gap of 100 bp
32425 33116: contig of 692 bp in length
33117 33216: gap of 100 bp
33217 33936: contig of 720 bp in length
33937 34036: gap of 100 bp
34037 34750: contig of 714 bp in length
34751 34850: gap of 100 bp
34851 35578: contig of 728 bp in length
35579 35678: gap of 100 bp
35679 36393: contig of 715 bp in length
36394 36493: gap of 100 bp
36494 37203: contig of 710 bp in length
37204 37303: gap of 100 bp
37304 37965: contig of 662 bp in length
37966 38065: gap of 100 bp
38066 38781: contig of 716 bp in length
38782 38881: gap of 100 bp
38882 39590: contig of 709 bp in length
39591 39690: gap of 100 bp
39691 40403: contig of 713 bp in length
40404 40503: gap of 100 bp
40504 41238: contig of 735 bp in length
41239 41338: gap of 100 bp

```
* 41339 42064: contig of 726 bp in length
* 42065 42164: gap of 100 bp
* 42165 42880: contig of 716 bp in length
* 42881 42980: gap of 100 bp
* 42981 43687: contig of 707 bp in length
* 43688 43787: gap of 100 bp
* 43788 44504: contig of 717 bp in length
* 44505 44604: gap of 100 bp
* 44605 45307: contig of 703 bp in length
* 45308 45407: gap of 100 bp
* 45408 46106: contig of 699 bp in length
* 46107 46206: gap of 100 bp
* 46207 46909: contig of 703 bp in length
* 46910 47009: gap of 100 bp
* 47010 47696: contig of 687 bp in length
* 47697 47796: gap of 100 bp
* 47797 48531: contig of 735 bp in length
* 48532 48631: gap of 100 bp
* 48632 49341: contig of 710 bp in length
* 49342 49441: gap of 100 bp
* 49442 50162: contig of 721 bp in length
* 50163 50262: gap of 100 bp
* 50263 50990: contig of 728 bp in length
* 50991 51090: gap of 100 bp
* 51091 51803: contig of 713 bp in length
* 51804 51903: gap of 100 bp
* 51904 52610: contig of 707 bp in length
* 52611 52710: gap of 100 bp
* 52711 53412: contig of 702 bp in length
* 53413 53512: gap of 100 bp
* 53513 54253: contig of 741 bp in length
* 54254 54353: gap of 100 bp
* 54354 55084: contig of 731 bp in length
* 55085 55184: gap of 100 bp
* 55185 55972: contig of 788 bp in length
* 55973 56072: gap of 100 bp
* 56073 56855: contig of 783 bp in length
* 56856 56955: gap of 100 bp
* 56956 57708: contig of 753 bp in length
* 57709 57808: gap of 100 bp
* 57809 58579: contig of 771 bp in length
* 58580 58679: gap of 100 bp

Query Match 19.8%; Score 226; DB 2; Length 214212;
Best Local Similarity 100.0%; Pred. No. 6.8e-118;
Matches 226; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 916 tctttcaattgtggaacagctccaagagcttgagtaggagagagtgcccgagggcg 975
Db 57085 TCTTTCAATTGTGGACAGCTCCAGAGGCTTGAGTAGGAGGAGTGCCCGGAGGCG 57144

QY 976 gggcgggcgggcggtgagctggctggcagtgggcggtggcggtgctccaggtgagc 1035
Db 57145 GGGCGGGCGGGCGGTGAGCTGGGCTGGCAGTGCGCGGTGCTGCCAGGTGAGC 57204

QY 1036 caccgtgcttctgccacagcagctcgctccacatccagctcttggctcctcgctt 1095
Db 57205 CACCGCTGCTTCTGCCACAGACAGGTGCGCTCCATCCAGGTCCTTGTGCTCGCTT 57264

QY 1096 gctgttctctttccacgcattttccaggataaactgtgactccagg 1141
Db 57265 GCCTGTCTCTTTTCCAGCATTTTCCAGGATAACTGTGACTCCAGG 57310

RESULT 6
BC020713
LOCUS Homo sapiens, Similar to serine (or cysteine) proteinase inhibitor,
DEFINITION clade B (ovalbumin), member 5, clone MGC:22489 IMAGE:4248249, mRNA,
complete cds.
ACCESSION BC020713
VERSION BC020713.1 GI:18089113
KEYWORDS MGC.
```

```
human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1289)
AUTHORS Strausberg, R.
JOURNAL Direct Submission
TITLE Submitted (03-JAN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONTECH
CDNA Library Preparation: CLONTECH Laboratories, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 35 Row: c Column: 23
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 4505788.
FEATURES
source
Location/Qualifiers
1..1289
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:22489 IMAGE:4248249"
/tissue_type="Prostate"
/clone_lib="NIH_MGC_83"
/lab_host="DH10B"
/note="vector: pDNR-LIB"
96..791
/codon_start=1
/product="Similar to serine (or cysteine) proteinase
inhibitor, clade B (ovalbumin), member 5"
/db_xref="GI:18089114"
/translation="MDALQLANSFAVDLFKOLCEKPLGNVLFSPICLSTLSLAOV
GAKGDTANEIGVLHFFENVKDVPFGFOTVTSQVANKLSFSYSLKLIKRLYVKSLSNLS
EFISSTRPYAKELETVDKDLBETKQINNSIKDLTDGHPENILADNSVNDQTKIL
VVNAAYFVGKWKMKFPPESETKECPFRVNVKVCAGCSSRRSPIIDVKNDNRDRVGHKSIP
MRNLRRAPAKLS"
BASE COUNT 401 a 274 c 280 g 334 t
ORIGIN

Query Match 7.8%; Score 89; DB 9; Length 1289;
Best Local Similarity 100.0%; Pred. No. 2e-39;
Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1053 agaacacgtgctccacatccaggtcttggctcctcgtgctgctgtcttccac 1112
Db 1 AGACACGGTCGCTCCACATCCAGGTCCTTGTGCTCCTCGCTGCTGCTTCTCTTCCAC 60

QY 1113 gcatttccaggataactgtgactccagg 1141
Db 61 GCATTTTCCAGGATAACTGTGACTCCAGG 89

RESULT 7
AX334331
LOCUS AX334331
DEFINITION Sequence
ACCESSION AX334331
VERSION AX334331.1 GI:18125050
KEYWORDS
```

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS 1 (sites)
TITLE Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G., Horrigan, S., Soppet, D.R. and Weaver, Z.
JOURNAL Cancer gene determination and therapeutic screening using signature gene sets
FEATURES Patent: WO 0194629-A 4840 13-DEC-2001; Avalon Pharmaceuticals (US)
source Location/Qualifiers
1..2566
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 786 a 526 c 504 g 750 t
ORIGIN

Query Match 5.3%; Score 61; DB 6; Length 2566;
Best Local Similarity 100.0%; Pred. No. 2.5e-23;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1081 ttgtgctcctgctgctgtctcttccacgcattttccaggataactgtgactccag 1140
|||||
Db 9 TTGTGCTCCTCGCTGCTGCTCTTCCACGCATTTCCAGGATAACTGTGACTCCAG 68
|||||

Qy 1141 g 1141
Db 69 G 69

RESULT 8
HSU04313
LOCUS HSU04313 2566 bp mRNA linear PRI 11-JUN-1994
DEFINITION Human maspin mRNA, complete cds.
ACCESSION U04313
VERSION U04313.1 GI:453368
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS Zou, Z., Anisowicz, A., Neveu, M., Rafidi, K., Sheng, S., Sager, R., Hendrix, M.J., Seftor, E. and Thor, A.
TITLE Maspin, a serpin with tumor suppressing activity in human mammary epithelial cells
JOURNAL Science 263, 526-529 (1994)
MEDLINE 94120413
REFERENCE Anisowicz, A.
AUTHORS Direct Submission
TITLE Submitted (10-DEC-1993) Anthony Anisowicz, Cancer Genetics, Dana Farber Cancer Institute, 44 Binney Street, Boston, MA 02115, USA
JOURNAL
FEATURES Location/Qualifiers
source 1..2566
/organism="Homo sapiens"
/isolate="76N"
/db_xref="taxon:9606"
/clone="232-1"
/sex="female"
/cell_line="76N"
/cell_type="epithelial"
/tissue_type="mammary"
/clone_lib="76N cDNA library in lambda Zap II"
1..75
76..1203
/codon_start=1
/product="maspin"
/protein_id="AA18957.1"
/db_xref="GI:453369"
/translation="MDALQLANSAFAVDLFKQLCEKPEPLGNVLFSPICLSTLSLAQV"

GAKGDTANEIGQVLFHFNVDKIDPFQGTVTSDVNVKLSYSYSLKLRRLVYDKSLNST
EFISSTKRPYAKELTVDFKDKLEETKQINNSIKDLTDGHFENILADNSVNDQTKIL
VNAAYEVGKMMKFPSETKCPFRNLKNTDTPQVMNMMEATFCMGNIDNSINCKLIE
LPQNKHLSPFILLPKDVEDFSGLEKIEKQLNESLSQWNPSTWNAKVKLSIKPKF
KVKMIDPKACLENLGLKHIFSEDSTDFSGMSETKGVALSNIHKVLCLETTEDGGDSI
EVPGARILQHKDELNADHPFIYIIRHNKTRNIIFFGKFCSP"
misc_feature 1093..1098
/note="putative serpin reactive center"
3'UTR 1201..2566
polyA_signal 2545..2550
BASE COUNT 786 a 526 c 504 g 750 t
ORIGIN

Query Match 5.3%; Score 61; DB 9; Length 2566;
Best Local Similarity 100.0%; Pred. No. 2.5e-23;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1081 ttgtgctcctgctgctgtctcttccacgcattttccaggataactgtgactccag 1140
|||||
Db 9 TTGTGCTCCTCGCTGCTGCTCTTCCACGCATTTCCAGGATAACTGTGACTCCAG 68
|||||

Qy 1141 g 1141
Db 69 G 69

RESULT 9
AR036974
LOCUS AR036974 2584 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 1 from patent US 5801001.
ACCESSION AR036974
VERSION AR036974.1 GI:5954830
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2584)
AUTHORS Sager, R., Zou, Z. and Anisowicz, A.
TITLE Method of detecting cancer
JOURNAL Patent: US 5801001-A 1 01-SEP-1998;
FEATURES Location/Qualifiers
source 1..2584
/organism="unknown"
BASE COUNT 804 a 526 c 504 g 750 t
ORIGIN

Query Match 5.3%; Score 61; DB 6; Length 2584;
Best Local Similarity 100.0%; Pred. No. 2.5e-23;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1081 ttgtgctcctgctgctgtctcttccacgcattttccaggataactgtgactccag 1140
|||||
Db 9 TTGTGCTCCTCGCTGCTGCTCTTCCACGCATTTCCAGGATAACTGTGACTCCAG 68
|||||

Qy 1141 g 1141
Db 69 G 69

RESULT 10
AR070462
LOCUS AR070462 2584 bp DNA linear PAT 18-FEB-2000
DEFINITION Sequence 1 from patent US 5905023.
ACCESSION AR070462
VERSION AR070462.1 GI:7221350
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2584)
AUTHORS Sager, R., Zou, Z. and Anisowicz, A.

TITLE Maspin, a serpin with tumor suppressing activity
JOURNAL Patent: US 5905023-A 1 18-MAY-1999;
FEATURES Location/Qualifiers
source 1..2584
/organism="unknown"

BASE COUNT 804 a 526 c 504 g 750 t
ORIGIN

Query Match 5.3%; Score 61; DB 6; Length 2584;
Best Local Similarity 100.0%; Pred. No. 2.5e-23;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1081 ttgtgctcctgcctgttccttttccacgcattttccaggataactgtgactccag 1140
|||||
Db 9 TTGTGCTCCTCGCTTCCCTGTTCTTTCCACGCATTTTCCAGGATAACTGTGACTCCAG 68

QY 1141 g 1141
|
Db 69 G 69

RESULT 11
LOCUS I15826 2584 bp DNA linear PAT 02-APR-1996
DEFINITION Sequence 1 from patent US 5470970.
ACCESSION I15826
VERSION I15826.1 GI:1250734
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 2584)
AUTHORS Sager, R., Anisowicz, A. and Zou, Z.
TITLE Maspin, a serpin with tumor suppressing activity
JOURNAL Patent: US 5470970-A 1 28-NOV-1995;
FEATURES Location/Qualifiers
source 1..2584
/organism="unknown"

BASE COUNT 804 a 526 c 504 g 750 t
ORIGIN

Query Match 5.3%; Score 61; DB 6; Length 2584;
Best Local Similarity 100.0%; Pred. No. 2.5e-23;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1081 ttgtgctcctgcctgttccttttccacgcattttccaggataactgtgactccag 1140
|||||
Db 9 TTGTGCTCCTCGCTTCCCTGTTCTTTCCACGCATTTTCCAGGATAACTGTGACTCCAG 68

QY 1141 g 1141
|
Db 69 G 69

RESULT 12
LOCUS I74315 2584 bp DNA linear PAT 03-APR-1998
DEFINITION Sequence 13 from patent US 5688641.
ACCESSION I74315
VERSION I74315.1 GI:3010456
KEYWORDS Unknown.
SOURCE Unknown.

ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 2584)
AUTHORS Sager, R., Zou, Z., Lee, S. Whan, and Tomasetto, C. Laure.
TITLE Cancer diagnosis using nucleic acid hybridization
JOURNAL Patent: US 5688641-A 13 18-NOV-1997;
FEATURES Location/Qualifiers
source 1..2584
/organism="unknown"

BASE COUNT 804 a 526 c 504 g 750 t
ORIGIN

Query Match 5.3%; Score 61; DB 6; Length 2584;
Best Local Similarity 100.0%; Pred. No. 2.5e-23;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1081 ttgtgctcctgcctgttccttttccacgcattttccaggataactgtgactccag 1140
|||||
Db 9 TTGTGCTCCTCGCTTCCCTGTTCTTTCCACGCATTTTCCAGGATAACTGTGACTCCAG 68

QY 1141 g 1141
|
Db 69 G 69

RESULT 13
LOCUS AL365181/1/c 164512 bp DNA linear PRI 22-JAN-2002
DEFINITION Human DNA sequence from clone RP11-284F21 on chromosome 1, complete sequence.

ACCESSION AL365181
VERSION AL365181.24 GI:18307339
KEYWORDS HTG.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
Chapman, J.
DIRECT SUBMISSION

REFERENCE 1 (bases 1 to 164512)
AUTHORS
TITLE
JOURNAL

COMMENT Submitted (22-JAN-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Jan 23, 2002 this sequence version replaced gi:18250487.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep
This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
RP11-284F21 is from the library RPC1-11.1 constructed by the group of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-284F21. It may be shorter because we sequence overlapping sections only once, except for a short overlap.
The true left end of clone RP11-284F21 is at 1 in this sequence.
The true left end of clone RP11-66D17 is at 162513 in this sequence. The true right end of clone RP11-98G7 is at 115430 in this sequence.

FEATURES Location/Qualifiers
source 1..164512
/organism="Homo sapiens"
/db_xref="taxon:9606"

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

/chromosome="1"
/clone="RP11-284F21"
/clone_lib="RPC1-11.1"
15086..15248
/note="Sequence from overlapping clone
RP11-98G7(AL139412). Assembly confirmed by restriction
digest."
16250..16282
/note="Sequence from overlapping clone
RP11-98G7(AL139412). Assembly confirmed by restriction
digest."
22737..22800
/note="Sequence from overlapping clone RP11-98G7
(AL139412). Assembly confirmed by restriction digest."
23982..24251
/note="Sequence from overlapping clone
RP11-98G7(AL139412). Assembly confirmed by restriction
digest."
25896..26069
/note="Sequence from overlapping clone RP11-98G7
(AL139412). Assembly confirmed by restriction digest."
29278..29453
/note="Sequence from overlapping clone
RP11-98G7(AL139412). Assembly confirmed by restriction
digest."
34394..34463
/note="Sequence from overlapping clone RP11-98G7
(AL139412). Assembly confirmed by restriction digest."
68335..69133
/note="Single clone region. Reads generated from a
transposon library derived from a single pUC clone.
Restriction digest data confirm the assembly."
68723..68836
/note="Sequence from uni-directional dGTP big dye
terminator reads only."
80361..80438
/note="Sequence from overlapping clone RP11-98G7
(AL139412). Assembly confirmed by restriction digest."
99673..99741
/note="Sequence from overlapping clone RP11-98G7
(AL139412). Assembly confirmed by restriction digest."
110040..110066
/note="Sequence from overlapping clone RP11-98G7
(AL139412). Assembly confirmed by restriction digest."
125123..125160
/note="Single clone region. Reads generated from a
transposon library derived from a single pUC clone.
Restriction digest data confirm the assembly."
BASE COUNT 41580 a 41780 c 40921 g 40231 t

Query Match 5.3%; Score 60; DB 9; Length 164512;
Best Local Similarity 100.0%; Pred. No. 1.4e-22;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 327 tttttagatgggttcacatgttccaggttgggttcacaaactcctgggctcaa 386
|||||
Db 100592 TTTTGTAGAGAGGGGCTCACCAGTTGCCAGGTGGTCTCAAACTCCTGGGCTCAA 100533

RESULT 14
AC013787/c 179307 bp DNA linear HTG 28-DEC-2001
LOCUS
DEFINITION Homo sapiens chromosome 15 clone RP11-531E13 map 15, *** SEQUENCING
IN PROGRESS ***, 9 unordered pieces.
AC013787 GI:17985927
VERSION HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 179307)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 15, clone RP11-531E13
Unpublished
2 (bases 1 to 179307)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Balducci,B., Barna,N., Beckerly,R., Boguslavsky,L., Boukhalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooper,P., DeArelano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lohocsky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meidrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (15-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Dec 28, 2001 this sequence version replaced gi:14190734.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3840
Center clone name: 531_E_13

* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 68483: contig of 68483 bp in length
* 68484 68583: gap of 100 bp
* 68584 75863: contig of 7280 bp in length
* 75864 75963: gap of 100 bp
* 75964 106600: contig of 30637 bp in length
* 106601 106700: gap of 100 bp
* 106701 120099: contig of 13399 bp in length
* 120100 120199: gap of 100 bp
* 120200 131661: contig of 11462 bp in length
* 131662 131761: gap of 100 bp
* 131762 134040: contig of 2279 bp in length
* 134041 134140: gap of 100 bp
* 134141 157957: contig of 23817 bp in length
* 157958 158057: gap of 100 bp
* 158058 177337: contig of 19279 bp in length
* 177337 177436: gap of 100 bp
* 177437 179307: contig of 1871 bp in length.
Location/Qualifiers
1. 179307
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="RP11-531E13"
/clone_lib="RPC1-11 Human Male BAC"
BASE COUNT 48710 a 40689 c 40017 g 49034 t 857 others
ORIGIN

Query Match 5.0%; Score 57; DB 2; Length 179307;


```

Best Local Similarity 100.0%; Pred. No. 7.4e-21;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 325 attttttgagagatgggtctcacatgttcccgaggttggtctcctcaactcctggg 381
|||||
Db 57986 ATTTTGTAGAGATGGGGCTCACCATGTTGCCAGGTGGTCTCAAACTCCTGGG 57930

RESULT 15
AC013391
LOCUS Homo sapiens chromosome 15 clone RP11-429B14 map 15, *** SEQUENCING
DEFINITION Homo sapiens chromosome 15 clone RP11-429B14 map 15, *** SEQUENCING
IN PROGRESS ***, 8 unordered pieces.
ACCESSION AC013391
VERSION AC013391.6 GI:16303477
KEYWORDS HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 208777)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 15, clone RP11-429B14
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 208777)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,W., Donelan,L., Doyle,M.,
Ferrelira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., MacDonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye.W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (09-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Oct 20, 2001 this sequence version replaced gi:12643085.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3803
Center clone name: 429_B_14
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 281: contig of 281 bp in length
* 282 381: gap of 100 bp
* 382 12491: contig of 12110 bp in length
* 12492 12591: gap of 100 bp
* 12592 14903: contig of 2312 bp in length
* 14904 15003: gap of 100 bp
* 15004 20280: contig of 5277 bp in length
* 20281 20380: gap of 100 bp
* 20381 27441: contig of 7061 bp in length
* 27442 27541: gap of 100 bp

```

```

* 27542 42474: contig of 14933 bp in length
* 42475 42574: gap of 100 bp
* 42575 119846: contig of 77272 bp in length
* 119847 119946: gap of 100 bp
* 119947 208777: contig of 88831 bp in length.
FEATURES
    location/Qualifiers
    source
    1..208777
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="15"
    /map="15"
    /clone="RP11-429B14"
    /clone_lib="RPC1-11 Human Male BAC"
BASE COUNT 54332 a 50627 c 49063 g 53962 t 793 others
ORIGIN
Query Match 5.0%; Score 57; DB 2; Length 208777;
Best Local Similarity 100.0%; Pred. No. 7.5e-21;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 325 attttttgagagatgggtctcacatgttcccgaggttggtctcctcaactcctggg 381
|||||
Db 8507 ATTTTGTAGAGATGGGGCTCACCATGTTGCCAGGTGGTCTCAAACTCCTGGG 8563

Search completed: September 24, 2002, 16:50:27
Job time: 10687 sec

```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2002, 14:45:25 ; Search time 197.2 Seconds
(without alignments)
9934.074 Million cell updates/sec

Title: US-09-617-174B-1

Perfect score: 1141

Sequence: 1 agataagcacagcagagaag.....agataactgtactccagg 1141

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_032802.*

- 1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
- 11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
- 12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
- 13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
- 14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
- 15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
- 16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
- 17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
- 18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
- 19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*
- 20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
- 21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
- 22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
- 23: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*
- 24: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1039	91.1	1141	18	AAT89000 Human maspin promo
2	89	7.8	410	21	AAC00200 Human secreted pro
3	61	5.3	2584	15	AAQ56790 Maspin DNA. Homo
4	61	5.3	2584	17	AAT07120 Tumour suppressor
5	61	5.3	2584	19	AAV55017 Maspin coding sequ
6	61	5.3	2584	20	AAAX36183 cDNA encoding huma
7	53	4.6	2136	22	AAH47053 Human maspin cDNA
8	49	4.3	12426	22	AAAS46482 Tumour suppressor
9	47	4.1	51	18	AAT89002 Human maspin oligo

10	43	3.8	14175	22	AAS27814	DNA encoding novel
11	43	3.8	14175	22	AAK78859	Human immune/haema
12	43	3.8	23164	22	AAK79678	Human immune/haema
13	43	3.8	26995	22	AAS29963	Human lung antigen
14	40	3.5	4086	21	AAA16677	Human secreted pro
15	40	3.5	12834	22	AAK73037	Human immune/haema
16	39	3.4	32191	22	AAS30497	DNA encoding novel
17	39	3.4	32191	22	AAAL06277	Human immune/haema
18	39	3.4	46107	22	AAK71730	Human immune/haema
19	38	3.3	8418	22	AAS41906	Genomic sequence #
20	38	3.3	10458	22	AAK77961	Human immune/haema
21	38	3.3	17874	22	AAK71070	Human immune/haema
22	38	3.3	33923	22	AAK67071	Human immune/haema
23	38	3.3	49999	20	AAZ23904	Human LOBO homolog
24	37	3.2	1388	22	AAK63064	Human immune/haema
25	37	3.2	1389	22	AAK69065	Human immune/haema
26	37	3.2	4505	22	AAK64819	Human immune/haema
27	37	3.2	9234	22	AAK89255	Human digestive sy
28	37	3.2	14150	22	AAK64820	Human immune/haema
29	37	3.2	465217	24	ABA90193	Human oestrogen re
30	36	3.2	534	22	AAH09878	Human cDNA clone (
31	36	3.2	1343	22	AAK67011	Human immune/haema
32	36	3.2	1705	22	AAH17516	Human cDNA sequenc
33	36	3.2	1717	22	AAH13941	Human cDNA sequenc
34	35	3.1	288	22	AAI86011	Human polynucleoti
35	35	3.1	330	22	AAI85941	Human polynucleoti
36	35	3.1	3303	16	AAQ76024	TGF-beta-3 promote
37	35	3.1	3303	22	AAH26121	Human transforming
38	35	3.1	22700	22	AAK70746	Human immune/haema
39	34	3.0	273	22	AAK64746	Human immune/haema
40	34	3.0	273	22	AAK64748	Human immune/haema
41	34	3.0	273	22	AAK64752	Human immune/haema
42	34	3.0	274	22	AAK64741	Human immune/haema
43	34	3.0	274	22	AAK64744	Human immune/haema
44	34	3.0	379	22	AAK67015	Novel human polynu
45	34	3.0	745	22	AAAL24371	Human breast cance

ALIGNMENTS

RESULT 1

AAT89000 AAT89000 standard; cDNA; 1141 BP.

ID AAT89000;

AC AAT89000;

XX 22-APR-1998 (first entry)

DE Human maspin promoter and partial coding cDNA.

XX Maspin; serpin; mammary epithelial cell; human; promoter; malignant;

KW tumour cell; treatment; prostate cancer; breast cancer; metastasis; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

FT promoter 1..1141

FT /tag= a

FT /note= "maspin promoter region"

FT misc_signal 443..449

FT /tag= b

FT misc_signal 451..457

FT /note= "Putative regulatory element AP2"

FT /tag= c

FT misc_signal 485..490

FT /note= "Putative regulatory element API"

FT /tag= d

FT misc_signal 684..691

FT /note= "Putative regulatory element Ets"

FT /tag= e

FT misc_signal 846..851

FT /note= "Putative regulatory element HRE"

FT /*tag= f
FT /Note= Putative regulatory element Ets"
FT 907..913
FT /*tag= g
FT /Note= Putative regulatory element APl"
FT prim_transcript 958..1141
FT /*tag= h
FT /Note= "partial"
XX
PN WO9736179-A1.
XX
XX 02-OCT-1997.
XX
XX 28-MAR-1997; 97WO-US05186.
XX
XX 28-MAR-1996; 96US-0014368.
XX
XX (DAND) DANA FARBER CANCER INST INC.
XX (DAND) DANA FARBER CANCER INST.
XX (PARD/) PARDEE A.
XX
XX Sagar R, Zhang M;
XX WPI; 1997-489785/45.
XX
XX Maspin gene promoter fragment - used to identify compounds for
XX treatment of prostate or breast cancer
XX
XX Claim 4; Fig 3; 51pp; English.

This sequence encodes the human maspin promoter region, including the
transcription start site. Maspin is a serpin which is expressed in
mammary epithelial cells. Its expression in these cells decreases
with increasing malignancy and is lost in during metastasis. Maspin
protein is also known to inhibit the mobility of tumour cells. This gene
can be used in method for screening compounds to identify candidate
compounds for the treatment of prostate cancer, or breast cancer. It can
also be used to identify compounds that increase the expression of
maspin, and for detecting the presence of metastatic prostate epithelial
cells.

XX Sequence 1141 BP; 261 A; 292 C; 277 G; 311 T; 0 other;

Query Match 91.1%; Score 1039; DB 18; Length 1141;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1139; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 agataagcacagacagagaaacacagctcggtttcaggtctcttctcaggtgattcg 60
DB 1 agataagcacagacagagaaacacagctcggtttcaggtctcttctcaggtgattcg 60
QY 61 gctgaaggagtaggttccaccacaaatgaagaagctgtggaagacagagagacaagaac 120
DB 61 gctgaaggagtaggttccaccacaaatgaagaagctgtggaagacagagagacaagaac 120
QY 121 aggctccagacagagatttcagagcagagctgcgtactcctttcttcttcttcttctt 180
DB 121 aggctccagacagagatttcagagcagagctgcgtactcctttcttcttcttcttctt 180
QY 181 gctctgtaccacaggtgaagtagctggttagctcaggtcactgcagcttcttgacctc 240
DB 181 gctctgtaccacaggtgaagtagctggttagctcaggtcactgcagcttcttgacctc 240
QY 241 ccaggctcaagtgtactctctcgtctcagctttcccaagtaactgggaccacaggtgcgt 300
DB 241 ccaggctcaagtgtactctcgtctcagctttcccaagtaactgggaccacaggtgcgt 300
QY 301 caccacgttaggctattgttttacatttttttagagatggggtctcaccattgttgccca 360
DB 301 caccacgttaggctattgttttacatttttttagagatggggtctcaccattgttgccca 360
QY 361 ggttggtctcaaaactcctgggctcagacgaatccgctcagctcaaacctcccaaatgctgg 420

DB 361 ggttggtctcaaaactcctgggctcaagcaatccgctcaacgctcaacccctcccaaatgctgg 420
QY 421 gattacaggtgagccacggccagggctgagtaataccttaatacacagagatttataaaa 480
DB 421 gattacaggtgagccacggccagggctgagtaataccttaatacacagagatttataaaa 480
QY 481 gaaactctcgcgcacccattaaacaatatctcctaccgaatttggttagtaataattttg 540
DB 481 gaaactctcgcgcacccattaaacaatatctcctaccgaatttggttagtaataattttg 540
QY 600 ctaatagtagtaataatttttagtaggcaactgtttatcacatatatccattcctctttt 600
DB 600 ctaatagtagtaataatttttagtaggcaactgtttatcacatatatccattcctctttt 600
QY 660 ttgattgtcttctgtttaatgggagctacctctcttggcatcttagcagaatgagctgc 660
DB 660 ttgattgtcttctgtttaatgggagctacctctcttggcatcttagcagaatgagctgc 660
QY 720 tgcagtttacacaaaagaatggagatcagagtagcttttttgcacacacagtgctgag 720
DB 720 tgcagtttacacaaaagaatggagatcagagtagcttttttgcacacacagtgctgag 720
QY 780 aaattttagttagttactatcacacattacttttttttttttttttttttttttttttt 780
DB 780 aaattttagttagttactatcacacattacttttttttttttttttttttttttttttt 780
QY 840 ggtcctgctggggccgagagagattgcccgaacgcatgtctgtacgtatgcatgtaactcac 840
DB 840 ggtcctgctggggccgagagagattgcccgaacgcatgtctgtacgtatgcatgtaactcac 840
QY 900 agccctctcctcccgaaacatgttgaggagccttttggagctgtgcagacacagacaact 900
DB 900 agccctctcctcccgaaacatgttgaggagccttttggagctgtgcagacacagacaact 900
QY 960 tcagctgaatcatctcttcaattgtgacaagctgccaagagcttgagtaggagagg 960
DB 960 tcagctgaatcatctcttcaattgtgacaagctgccaagagcttgagtaggagagg 960
QY 1020 agtgcgcgcagcgg 1020
DB 961 agtgcgcgcagcgg 1020
QY 1080 gctgccaggtgagccacgcgtctcttctgccagacacgctgcgtccacatccaggtct 1080
DB 1021 gctgccaggtgagccacgcgtctcttctgccagacacgctgcgtccacatccaggtct 1080
QY 1081 ttgtgtcctcgttgcgttcttcttccacgcattttccaggataactgtgactccag 1140
DB 1081 ttgtgtcctcgttgcgttcttcttccacgcattttccaggataactgtgactccag 1140
QY 1141 g 1141
DB 1141 g 1141

RESULT 2

AAC00200
ID AAC00200 standard; cDNA; 410 BP.

XX AAC00200;

XX XX

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 198.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

KW gene therapy; chromosome mapping; ss.

XX Homo sapiens.

OS OS

XX XX

PN EPI033401-A2.

```

PD 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-0200610.
XX
XX 26-FEB-1999; 99US-0122487.
XX
XX (GEST ) GENSET.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
PI
PI WPI; 2000-500381/45.
DR
DR P-PSDB; AAG00194.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 1; SEQ ID 198; 71pp + CD-ROM; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.
XX
XX Sequence 410 BP; 103 A; 108 C; 76 G; 122 T; 1 other;
SQ
Query Match 7.8%; Score 89; DB 21; Length 410;
Best Local Similarity 100.0%; Pred. No. 2.9e-34;
Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1053 agacacggtcgctccacatccagctgttctgtctctctgtctgtctcttccac 1112
DB 1 agacacggtcgctccacatccagctgttctgtctctctgtctgtctcttccac 60
QY 1113 gcattttccagataactgtgactccag 1141
DB 61 gcattttccagataactgtgactccag 89
RESULT 3
AAQ56790
ID AAQ56790 standard; DNA; 2584 BP.
XX
XX AAQ56790;
XX
XX 30-SEP-1994 (first entry)
DE Maspin DNA.
XX
XX Maspin; serine protease inhibitor; mammary; epithelial cells; cancer;
KW breast tissue; well-differentiated; poorly differentiated; antibody;
KW primary carcinoma; lymph node lesion; metastatic lesion; carcinoma;
KW tumour; tumour suppression; invasion suppression; anti-maspin;
KW anticancer compound; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH 76..1203
FT CDS /*tag= a
FT /product= Maspin
XX

```

```

PN WO9405804-A.
XX
XX 17-MAR-1994.
XX
XX 01-SEP-1993; 93WO-US08322.
XX
XX 01-SEP-1992; 92US-0938823.
XX
XX (DAND ) DANA FARBER CANCER INST INC.
XX
XX Sager R;
PI
PI WPI; 1994-101208/12.
DR
DR P-PSDB; AAR50938.
XX
XX DNA encoding maspin which has tumour suppressing activity - used
PT to produce prods. for the diagnosis, staging and treatment of
PT carcinoma and for screening anti-cancer agents
XX
XX Claim 1; Page 35-37; 67pp; English.
XX
XX This sequence represents the maspin gene which was isolated by
CC subtractive hybridisation. This gene encodes a serine protease
CC inhibitor and is expressed in normal mammary epithelial cells in
CC culture and in normal breast tissue. Expression of this gene
CC decreases during progression from well-differentiated to poorly
CC differentiated primary carcinomas, and is absent in most lymph node
CC and distant metastatic lesions. The maspin protein has tumour
CC suppressing and invasion suppressing activity. The DNA and anti-
CC maspin antibodies can be used to determine whether test cells are
CC cancerous or for determining whether a test carcinoma cell represents
CC an early-stage or advanced metastatic carcinoma. The level of
CC expression of maspin in cells can be used for screening anticancer
CC compounds. Compounds which increase the amount of maspin in, or in
CC the immediate vicinity of carcinoma cells can be used to treat
CC carcinomas. Maspin, and DNA encoding maspin, can also be used in
CC the treatment of carcinomas
XX
XX Sequence 2584 BP; 804 A; 526 C; 504 G; 750 T; 0 other;
SQ
Query Match 5.3%; Score 61; DB 15; Length 2584;
Best Local Similarity 100.0%; Pred. No. 3.1e-20;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1081 ttgtgctcctgctgtctgtcttcttccacgcatcttccaggataactgtgactccag 1140
DB 9 ttgtgctcctgctgtctgtcttcttccacgcatcttccaggataactgtgactccag 68
QY 1141 g 1141
DB 69 g 69
RESULT 4
AAAT07120
ID AAAT07120 standard; cDNA; 2584 BP.
XX
XX AAAT07120;
XX
XX 15-FEB-1996 (first entry)
DE Tumour suppressor maspin cDNA.
XX
XX Maspin; serpin; tumour suppressor; breast cancer; carcinoma;
KW gene therapy; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH 76..1203
FT CDS /*tag= a
XX

```

Wed Sep 25 08:14:15 2002

us-09-617-174b-1.rng

```

PR 01-SEP-1993; 93US-0121714.
PR 28-FEB-1991; 91US-0662216.
PR 28-FEB-1992; 92US-0844296.
PR 01-SEP-1992; 92US-0938823.
PR 07-JUN-1995; 95US-0477108.
XX
XX (DAND ) DANA FARBER CANCER INST INC.
XX
XX Anisowicz A, Sager R, Zou Z;
XX
XX WPI; 1998-494765/42.
XX P-PSDB; AAW69287.
XX
XX Detection of maspin, a serine protease inhibitor - using antibodies
XX raised against it, in the immuno-diagnosis of carcinoma(s)
XX associated with lack of tumour suppression
XX
XX PS Disclosure; Fig 3; 31pp; English.
XX
XX This sequence encodes the human maspin protein, which is a member of the
XX serine protease inhibitor (serpin) family. The protein can be detected
XX using the method of the invention, which comprises: (a) providing a test
XX cell from a given type of epithelial tissue, suspected of being
XX cancerous; (b) contacting proteins of the test cell with an antibody (Ab)
XX which forms an immunocomplex with polypeptide maspin; and (c) comparing:
XX (i) the amount of immunocomplex formation by the Ab and the proteins of
XX (ii) the amount of immunocomplex formation by the Ab and the proteins of
XX a normal control cell from the type of epithelial tissue, where an amount
XX of immunocomplex formation with the proteins of the test cell
XX substantially less than the amount obtained with the proteins of the
XX normal cells is an indication that the test cell is cancerous. The method
XX is used for the detection of cancers, especially mammary epithelial
XX associated cancers. Maspin, is involved in tumour suppression. Metastases
XX and invasive tumours are characterised by increased proteolytic activity,
XX especially of serine proteases. Maspin is expressed in normal mammary
XX epithelial cells, and a decrease in expression is seen as primary
XX carcinomas progress. By using (monoclonal) Abs raised against it,
XX immunological based detection techniques e.g. ELISA, can be used to
XX screen for maspin levels in biological samples e.g. blood, serum, urine,
XX saliva, milk, ductal fluid, tears or semen. Reduced levels of maspin,
XX compared to normal, indicate lack of tumour suppression and thus the
XX presence of carcinomic cells.
XX
XX Sequence 2584 BP; 804 A; 526 C; 504 G; 750 T; 0 other;
XX
XX
XX Query Match 5.3%; Score 61; DB 19; Length 2584;
XX Best Local Similarity 100.0%; Pred. No. 3.1e-20;
XX Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1081 ttgtgctcgtgctgctgttcttccacgcattttccaggataactgtgactccag 1140
XX |
XX Db 9 ttgtgctcgtgctgctgttcttccacgcattttccaggataactgtgactccag 68
XX
XX QY 1141 g 1141
XX |
XX Db 69 g 69
XX
XX RESULT 6
XX AAX36183
XX ID AAX36183 standard; cDNA; 2584 BP.
XX
XX AC AAX36183;
XX
XX DT 19-JUL-1999 (first entry)
XX
XX XX cDNA encoding human maspin protein.
XX DE Human; serpin; maspin; growth inhibition; carcinoma cell;
XX KW genetic therapy; down-regulation; maspin expression; ds.
XX XX Homo sapiens.
XX OS
XX
XX US5470970-A.
XX
XX PD 28-NOV-1995.
XX
XX PF 28-FEB-1991; 91US-0662216.
XX
XX PR 01-SEP-1993; 93US-0121714.
XX PR 28-FEB-1991; 91US-0662216.
XX PR 28-FEB-1992; 92US-0844296.
XX PR 01-SEP-1992; 92US-0938823.
XX
XX (DAND ) DANA FARBER CANCER INST INC.
XX
XX Anisowicz A, Sager R, Zou Z;
XX
XX WPI; 1996-019941/02.
XX P-PSDB; AAR66882.
XX
XX Novel tumour suppressor protein Maspin - useful for treating
XX carcinoma(s), esp. breast cancer
XX
XX Claim 1: Column 19-24; 30pp; English.
XX
XX Tumour suppressor maspin cDNA (AA071120) was isolated from a normal
XX human mammary epithelial cell library (76N). The gene was mapped
XX to 18q21.3. Expression of maspin decreases during carcinoma
XX progression, and is absent in most lymph node and distant
XX metastatic lesions. The isolated DNA can be used to produce
XX recombinant maspin, and in diagnostic, screening and therapeutic
XX methods, including cancer gene therapy.
XX
XX Sequence 2584 BP; 804 A; 526 C; 504 G; 750 T; 0 other;
XX
XX
XX Query Match 5.3%; Score 61; DB 17; Length 2584;
XX Best Local Similarity 100.0%; Pred. No. 3.1e-20;
XX Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1081 ttgtgctcgtgctgctgttcttccacgcattttccaggataactgtgactccag 1140
XX |
XX Db 9 ttgtgctcgtgctgctgttcttccacgcattttccaggataactgtgactccag 68
XX
XX QY 1141 g 1141
XX |
XX Db 69 g 69
XX
XX RESULT 5
XX AAV55017
XX ID AAV55017 standard; DNA; 2584 BP.
XX
XX AC AAV55017;
XX
XX DT 11-NOV-1998 (first entry)
XX
XX DE Maspin coding sequence.
XX
XX KW Maspin; human; serine protease inhibitor; serpin; cancer detection;
XX KW mammary epithelial associated cancer; tumour suppressor; ds.
XX
XX OS Homo sapiens.
XX
XX XX Location/Qualifiers
XX FH 76.1203
XX FT CDS /*tag= a
XX FT /*product= Maspin
XX
XX US5801001-A.
XX
XX PD 01-SEP-1998.
XX
XX PF 07-JUN-1995; 95US-0477108.
XX
XX
```

```

XX PN US905023-A.
XX PD 18-MAY-1999.
XX PF 07-JUN-1995; 95US-0477112.
XX PR 01-SEP-1993; 93US-0121714.
XX PR 28-FEB-1991; 91US-0662216.
XX PR 28-FEB-1992; 92US-0844296.
XX PR 01-SEP-1992; 92US-0938823.
XX PR 07-JUN-1995; 95US-0477112.
XX PA (DAND ) DANA FARBER CANCER INST INC.
XX PI Anisowicz A, Sager R, Zou Z;
XX DR WPI; 1999-326381/27.
XX DR P-PSDB; AAY02607.
XX PT Inhibiting the growth of a carcinoma using a serpin with tumour
XX suppressing activity
XX PS Claim 1; Columns 17-22; 29pp; English.
XX CC The present sequence encodes a human serpin called maspin. The
XX specification describes a method for inhibiting the growth of
XX a carcinoma by treating carcinoma cells with maspin. The method
XX is useful as genetic therapy to help control carcinomas
XX characterized by down-regulated maspin expression.
XX SQ Sequence 2584 BP; 804 A; 526 C; 504 G; 750 T; 0 other;

Query Match 5.3%; Score 61; DB 20; Length 2584;
Best Local Similarity 100.0%; Pred. No. 3.1e-20;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1081 ttgtactcgtctgctgtctctttccacgcattttccaggataactgtgactccag 1140
Db 9 tttgtctctcgtctgctgtctctttccacgcattttccaggataactgtgactccag 68

Oy 1141 g 1141
Db 69 g 69

RESULT 7
AAH47053
ID AAH47053 standard; DNA; 2136 BP.
AC AAH47053;
XX
DT 29-OCT-2001 (first entry)
XX
DE Human maspin cDNA sequence.
XX
KW Drug resistance; resistance gene; semaphorin D; B94; mel-14 antigen;
KW 24p3; proliferin; maspin; cancer; cytostatic; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN WO200155455-A2.
XX
PD 02-AUG-2001.
XX
XX 31-JAN-2001; 2001WO-US03161.
XX
XX 31-JAN-2000; 2000US-0179191.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX (JINS/) JIN S.
XX

```

```

PI Jin S;
XX WPI; 2001-488799/53.
XX
XX Determining if a compound modulates the drug resistance of a cell,
XX comprises determining the expression or activity level of a resistance
XX sequence in a cell in the presence of the test compound -
XX
XX Example 1; Fig 5A-B; 79pp; English.
XX
XX The invention relates to a method of determining whether a test compound
XX modulates the drug resistance of a cell that comprises determining the
XX expression or activity level of resistance genes (e.g. semaphorin D, B94,
XX mel-14 antigen, 24p3, proliferin or maspin) in a cell in the presence of
XX the test compound, and comparing its expression or activity level in a
XX cell without the test compound. The drug resistant sequences are useful
XX in identifying drug resistant cells, in screening methods directed to the
XX identification of compounds that can modulate the drug resistance of a
XX cell type or multiple cell types. An isolated resistance protein can be
XX used as an immunogen to generate antibodies that bind the resistance
XX protein. Resistance nucleic acids may be inserted into vectors and used
XX as gene therapy vectors. An anti-resistance protein antibody may be used
XX to isolate a resistance protein, or facilitate the purification of
XX natural resistance protein from cells and of recombinantly produced
XX resistance protein expressed in host cells. The methods are useful for
XX treating a subject having a disorder, such as a drug-resistance cancer,
XX characterized by aberrant resistance sequence expression or activity by
XX administering to the subject a resistance modulator. The present sequence
XX represents a human maspin cDNA sequence, whose expression was decreased
XX in drug resistant EMT6 tumours.
XX SQ Sequence 2136 BP; 658 A; 454 C; 428 G; 596 T; 0 other;

Query Match 4.6%; Score 53; DB 22; Length 2136;
Best Local Similarity 100.0%; Pred. No. 3.2e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1089 ctccgttaccgttctctttccacgcattttccaggataactgtgactccag 1141
Db 87 ctccgttaccgttctctttccacgcattttccaggataactgtgactccag 139

RESULT 8
AAS46482/C
ID AAS46482 standard; DNA; 12426 BP.
XX
XX AC AAS46482;
XX
XX 18-DEC-2001 (first entry)
XX
XX Tumour suppressor gene derived chemically modified sequence #204.
XX
XX Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
XX cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
XX cytosine methylation; ds.
XX
XX OS Homo sapiens.
XX
XX PN WO200168912-A2.
XX
XX PD 20-SEP-2001.
XX
XX 15-MAR-2001; 2001WO-EP02955.
XX
XX 15-MAR-2000; 2000DE-1013847.
XX
XX 06-APR-2000; 2000DE-1019058.
XX
XX 07-APR-2000; 2000DE-1019173.
XX
XX 30-JUN-2000; 2000DE-1032529.
XX
XX 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIG-) EPIGENOMICS AG.
XX

```

PR	28-MAR-1996;	96US-0014368.
XX	(DAND)	DANA FARBER CANCER INST INC.
PA	(DAND)	DANA FARBER CANCER INST.
PA	(PARD/)	PARDEE A.
XX	Sagar R,	Zhang M;
PI		
XX	WPI;	1997-489785/45.
XX	Maspin gene promoter fragment - used to identify compounds for	
PT	treatment of prostate or breast cancer	
XX	Disclosure; Page 15; 51pp; English.	
PS		
XX	Primer OLI is designed from the 5' untranslated region of maspin and is	
CC	used screen genomic DNA to isolate the maspin upstream promoter sequence.	
CC	Maspin is a serpin which is expressed in mammary epithelial cells. Its	
CC	expression in these cells decreases with increasing malignancy and is	
CC	lost in during metastasis. Maspin protein is also known to inhibit the	
CC	mobility of tumour cells. This gene can be used in method for screening	
CC	compounds to identify candidate compounds for the treatment of prostate	
CC	cancer, or breast cancer. It can also be used to identify compounds that	
CC	increase the expression of maspin, and for detecting the presence of	
CC	metastatic prostate epithelial cells.	
XX		
Seq	Sequence 51 BP; 18 A; 10 C; 16 G; 7 T; 0 other;	
	Query Match	4.1%; Score 47; DB 18; Length 51;
	Best Local Similarity	100.0%; Pred. No. 2.9e-13;
	Matches 47; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Oy	1085	gctcctgcgttcgctgttccctttccacgcatttccaggataacty ll31
ph	51	GCTCCGCGTGCCTGTTCCTTTTCCACGCATTTTCCAGGATAACTG 5

RESULT	10	
AAS27814		
ID	AAS27814	standard; DNA; 14175 BP.
XX		
XX	AAS27814;	
XX		
XX		
DT	07-NOV-2001	(first entry)
XX		
DE	DNA	encoding novel signal transduction pathway protein, Seq ID 1474.
XX		
KW	Neuroprotective;	cytostatic; dermatological; immunosuppressive; tumour;
KW	antiinflammatory;	anti-HIV; antibacterial; antiinflammatory; cancer;
KW	immune system disorder;	rheumatoid arthritis; inflammatory condition;
KW	organ transplant rejection;	infection; hepatitis C; blood disorder;
KW	sickle cell anaemia;	hyperproliferative disorder; Gaucher's disease;
KW	neurodegenerative disorder;	Alzheimer's disease; Parkinson's disease;
KW	chromosomal abnormality;	Down syndrome; ischaemia; renal disorder;
KW	cardiovascular;	respiratory; wound healing; endocrine; Addison's disease;
KW	reproductive system;	gastrointestinal; liver disorder; AIDS; ds;
KW	acquired immune deficiency	syndrome.
XX		
OS	Homo sapiens.	
XX		
PN	WO200154733-A1.	
XX		
PD	02-AUG-2001.	
XX		
PF	17-JAN-2001;	2001WO-US01312.
XX		
XX	31-JAN-2000;	2000US-0179065.
PR	04-FEB-2000;	2000US-0180628.
PR	24-FEB-2000;	2000US-0184664.
PR	02-MAR-2000;	2000US-0186350.
PR	16-MAR-2000;	2000US-0189874.
PR	17-MAR-2000;	2000US-0190076.

PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225265.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 12-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.

PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251899.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465460/50.
XX
XX Novel polypeptides useful for diagnosing, treating, preventing and/or
PT prognosing disorders related to the proteins, including cancers, immune
PT disorders and neuronal disorders
XX
PS Claim 1; SEQ ID No 1474; 880pp; English.
XX
XX The invention relates to novel isolated polypeptides (I), and
CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for
CC diagnosing, preventing and treating diseases including immune system

CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ transplant rejections and graft versus host disease, infectious diseases (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and other blood-related disorders (sickle cell anaemia), myeloproliferative disorders, primary haematopoietic disorders, hyperproliferative disorders (e.g. Gaucher's disease and cancer), neurodegenerative disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal disorders (e.g. glomerulonephritis), cardiovascular disorders (e.g. arrhythmia), respiratory disorders, dermatological disorders, in wound healing, epithelial cell proliferation, endocrine disorders (e.g. Addison's disease), reproductive system disorders, gastrointestinal disorder (inflammatory disorders), liver disorders (cirrhosis), as stimulators of B-cell responsiveness to pathogens, activators of T-cells, to induce higher affinity antibodies, and as a means to induce tumour proliferation in pathologies e.g. acquired immune deficiency syndrome (AIDS). AAS26976-AAS27850 represent novel signal transduction pathway protein coding sequences and PCR primers of the invention.

XX Sequence 14175 BP; 3111 A; 3818 C; 4137 G; 3109 T; 0 other;
SQ Query Match 3.8%; Score 43; DB 22; Length 14175;
Best Local Similarity 100.0%; Pred. No. 3.4e-11;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 351 atgtgtccagggttgctcaactctggctggtcgaagcaatcc 393
|||||
Db 9809 atgtgtccagggttgctcaactctggctggtcgaagcaatcc 9851

RESULT 11
AAK78859/G
ID AAK78859 standard; DNA; 14175 BP.
XX
AC AAK78859;
XX
DT 07-NOV-2001 (first entry)
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:33671.
DE Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
KW Homo sapiens.
OS
XX WO200157182-A2.
XX
XX 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0226279.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.

PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI: 2001-483426/52.
XX

PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis.
XX
PS Disclosure; SEQ ID NO 33671; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins, and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX

SQL Sequence 14175 BP; 3109 A; 4137 C; 3818 G; 3111 T; 0 other;
Query Match 3.8%; Score 43; DB 22; Length 14175;
Best Local Similarity :00.0%; Pred. No. 3.4e-11;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 351 atgttgccagggttggtctcaaaactcctgggtcgaagcaatcc 393
|||||
DB 4367 ATGTTGCCAGGTGGTCTCAAACTCCTGGGCTCAAGCAATCC 4325
RESULT 12
AAK79678/c
ID AAK79678 standard; DNA; 23164 BP.
XX
AC AAK79678;
XX
DT 07-NOV-2001 (first entry)
XX Human immune/hematopoietic antigen genomic sequence SEQ ID NO:34490.
XX Human; immune; hematopoietic; immune/hematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
XX WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 23-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.

XX Human lung antigen genomic DNA #33.
DE
XX
KW Chicken; sheep; immunosuppressive; antiarthritic; vasotropic;
KW chicken; goat; rabbit; horse; cat; dog;
KW antineumatic; antiproliferative; cytostatic; cardiant; neuroprotective;
KW cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;
KW ophthalmological; vulnary; gene therapy; autoimmune disease; neoplasm;
KW hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;
KW cerebrovascular disorder; nervous system disorder; bacterial infection;
KW fungal infection; viral infection; ocular disorder; endocrine disorder;
KW gastrointestinal disorder; renal disorder; respiratory disorder;
KW wound healing; skin aging; organ transplantation; food preservative;
KW tissue regeneration; anti-infertility; food additive.
XX
OS Homo sapiens.
XX
XX WO200155303-A2.
PN
XX
PD
XX
XX 02-AUG-2001.
PF
XX 17-JAN-2001; 2001WO-US01301.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 26-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.

PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-457723/49.
XX
XX Isolated polypeptide for treating, preventing and/or prognosing
XX respiratory disorders related to the lung including lung cancers and
XX also for testing and detection e.g. diagnosis -
XX
XX Claim 1; SEQ ID NO 227; 507pp; English.
XX
XX Sequences AAS29931-AAS30164 represent genomic DNA molecules, which encode
XX the lung antigen polypeptides of the invention. Lung antigen polypeptides
XX and their associated polynucleotides are useful in the diagnosis,
XX treatment and prevention of various types of disorders in e.g. humans,
XX mice, rabbits, goats, horses, cats, dogs, chickens or sheep. A
XX pathological condition can be determined by detecting the presence or
XX absence of a mutation in a lung antigen polynucleotide. The treatable
XX disorders include autoimmune diseases such as rheumatoid arthritis,
XX hyperproliferative disorders such as neoplasms of the breast or liver,
XX cardiovascular disorders such as cardiac arrest, cerebrovascular
XX disorders such as cerebral ischaemia, nervous system disorders such as
XX Alzheimer's disease, infections caused by bacteria, viruses and fungi,
XX ocular disorders such as corneal infection, endocrine disorders such as
XX premature labour and infertility, gastrointestinal disorders such as
XX Crohn's disease, renal disorders such as glomerulonephritis and
XX respiratory disorders such as asthma and pleurisy. The polypeptides can
XX also be used to aid wound healing, to prevent skin aging due to sunburn,
XX to maintain organs before transplantation, to regenerate tissues and in
XX chemotaxis. The polypeptides can also be used as a food additive or
XX preservative to increase or decrease storage capabilities.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

Query Match 3.8%; Score 43; DB 22; Length 26995;
Best Local Similarity 100.0%; Pred. No. 3.4e-11;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 346 tcacatgttccagtggtgtctcaaacctctggctcaagc 388
|||||
DB 5866 TCACCATGTTGCCAGGTGGTCTCAAACTCTGGGCTCAAGC 5824

RESULT 14
AAAL6677
ID AAAL6677 standard; cDNA; 4086 BP.
XX
XX AAAL6677;
XX
XX 16-JUN-2000 (first entry)
XX
XX Human secreted protein clone kv10_8 nucleotide sequence SEQ ID NO:119.
XX
XX Human; secreted protein; immunestimulant; immunesuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;

KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX Homo sapiens.
XX WO200009552-A1.
XX
XX 24-FEB-2000.
XX
XX 13-AUG-1999; 99WO-US18298.
XX
XX 14-AUG-1998; 98US-0096622.
XX 17-AUG-1998; 98US-0096815.
XX 04-SEP-1998; 98US-0099229.
XX 23-OCT-1998; 98US-0105368.
XX 08-JAN-1999; 99US-0115234.
XX 12-FEB-1999; 99US-0119931.
XX 18-FEB-1999; 99US-0120575.
XX 30-APR-1999; 99US-0132020.
XX 11-AUG-1999; 99US-0096622.
XX (GENY) GENETICS INST INC.
XX
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
XX Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
XX Wong GG, Clark HF, Fechtel K;
XX WPI; 2000-205979/18.
XX P-PSDB; AAY94957.
XX
XX New polynucleotides encoding secreted proteins, which may have e.g.
XX nutritional, chemokine, immune stimulating or suppressing,
XX hematopoiesis regulating, tissue growth, activin/inhibin
XX antiinflammatory or tumor inhibition activity -
XX
XX Claim 128; Page 579-580; 641pp; English.
XX
XX AAAL6618 to AAAL6697 encode the human secreted proteins given in
XX AAY94898 to AAY94980, isolated from human adult brain, adult thyroid,
XX adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
XX adult placenta, adult testis, whole embryo, adult cartilage, kidney,
XX foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
XX and adult bladder, cDNA libraries. The polynucleotides and proteins are
XX predicted to have biological activities which would make them suitable
XX for treating, preventing or ameliorating medical conditions in humans
XX and animals. The polynucleotides can be used as markers for tissues in
XX which the protein is preferentially expressed, as molecular weight
XX markers on Southern gels; and as chromosome markers or tags to identify
XX chromosomes or to map gene positions. The proteins can be used in the
XX treatment of immune deficiencies and disorders, such as severe combined
XX immunodeficiency (SCID), as well as viral, bacterial, fungal and other
XX infections. These infections include human immunodeficiency virus (HIV),
XX hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
XX candidiasis. The proteins can be used to treat autoimmune disorders such
XX as connective tissue disease, multiple sclerosis, systemic lupus
XX erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
XX Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
XX diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
XX autoimmune inflammatory eye disease. The proteins can also be used to
XX treat allergic conditions, such as asthma. AAAL6698 to AAAL6774 represent
XX probes for the human secreted proteins from the present invention.
XX
XX Sequence 4086 BP; 1031 A; 942 C; 1014 G; 1099 T; 0 other;

Query Match 3.5%; Score 40; DB 21; Length 4086;
Best Local Similarity 100.0%; Pred. No. 1e-09;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 346 tcaccatgttgcacaggttggtctcaaaactcctcgtgggctca 395
|||||
Db 1982 tcaccatgttgcacaggttggtctcaaaactcctcgtgggctca 2021

RESULT 15
AAK73037
ID AAK73037 standard; DNA; 12834 BP.
XX
AC AAK73037;
DT 06-NOV-2001 (first entry)
XX
DE Human Immune/haematopoietic antigen genomic sequence SEQ ID NO:27849.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytotstatic; gene therapy; vaccine; metastasis; ds.
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.

PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2002, 13:46:30 ; Search time 1616.37 Seconds
(without alignments)
9527.537 Million cell updates/sec

Title: US-09-617-174B-1
Perfect score: 1141
Sequence: 1 agataagcacagagagaag.....aggataactgtgactccagg 1141

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	136	11.9	651	9	AL602439
2	97	8.5	738	10	BF208602
3	89	7.8	527	9	AL602677
4	89	7.8	999	10	BF676061
5	83	7.3	714	10	BF723255
6	59	5.2	442	10	H74001
7	55	4.8	412	10	H49484
8	55	4.8	648	12	AG100732
9	54	4.7	625	9	AL602554
10	47	4.1	288	9	AA382130
11	45	3.9	382	9	AA565270
12	45	3.9	395	9	AA713761
13	45	3.9	484	9	AA550845
14	45	3.9	931	10	BM008408
15	44	3.9	328	10	N53046
16	44	3.9	335	10	T96921
17	44	3.9	417	10	N73230

C 18	43	3.8	204	9	AW268077
C 19	43	3.8	248	9	AA494274
C 20	43	3.8	296	12	AQ089138
C 21	43	3.8	328	9	AA493880
C 22	41	3.6	382	12	AQ147846
C 23	40	3.5	154	9	AA309434
C 24	40	3.5	293	9	AW961841
C 25	40	3.5	391	9	AW851810
C 26	40	3.5	413	9	AL600265
C 27	39	3.4	304	9	AI803767
C 28	39	3.4	335	10	BI020761
C 29	39	3.4	340	10	BI020795
C 30	39	3.4	381	10	BF840682
C 31	39	3.4	381	10	BF925361
C 32	39	3.4	381	10	BF925367
C 33	39	3.4	381	10	BF925371
C 34	39	3.4	381	10	BF925374
C 35	39	3.4	382	10	BF925369
C 36	39	3.4	383	10	BF925368
C 37	39	3.4	401	12	AQ481517
C 38	39	3.4	416	10	BF896241
C 39	39	3.4	416	10	BF896244
C 40	39	3.4	418	10	BF896243
C 41	39	3.4	617	12	B64146
C 42	39	3.4	639	9	AA316156
C 43	38	3.3	338	9	AA503501
C 44	38	3.3	426	10	BF840283
C 45	38	3.3	443	9	AA720734
					AW268077 xr92d10.x
					AA494274 ng96b12.s
					AQ089138 HS-2207_B
					AA493880 nh07a07.s
					AQ147846 HS-3091_A
					AA309434 EST180303
					AW961841 EST373914
					AW851810 QVO-CT022
					AL600265 DKF2p3130
					AI803767 tp29h07.x
					BI020761 CM4-MT021
					BI020795 CM4-MT021
					BF840682 IL0-HT106
					BF925361 IL0-NT023
					BF925367 IL0-NT023
					BF925371 IL0-NT023
					BF925374 IL0-NT023
					BF925369 IL0-NT023
					BF925368 IL0-NT023
					AQ481517 RPC1-11-2
					BF896241 IL0-MT021
					BF896244 IL0-MT021
					BF896243 IL0-MT021
					B64146 CIT-HSP-202
					AA316156 EST187893
					AA503501 nh59h03.s
					BF840283 RC4-HT04.8
					AA720734 nv93d04.r

ALIGNMENTS

RESULT 1
AL602439 651 bp mRNA linear EST 14-AUG-2001
LOCUS DKFZp686L1213_r1 686 (synonym: hicc3) Homo sapiens EST clone
DEFINITION DKFZp686L1213 5', mRNA sequence.
ACCESSION AL602439
VERSION AL602439.1 GI:15165945
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 551)
AUTHORS Ansong, W., Wirkner, U., Mewes, H.W., Weil, B. and Wiemann, S.)
TITLE EST (Ansong, W., Wirkner, U., Mewes, H.W., Weil, B. and Wiemann, S.)
JOURNAL Unpublished (1999)
COMMENT Contact: Ansong W
MIPS

Am Klopferspitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by EMBL (European Molecular Biology Laboratories,
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project
No SI sequence available.
This clone (DKFZp686L1213) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rpd.de.

FEATURES

Location/Qualifiers
1..651
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp686L1213"
/clone.lib="686 (synonym: hicc3)"
/tissue.type="human skeletal muscle"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pTriplex2; Site_1: SflIA; Site_2: SflIB;
cDNA-collection"

By

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 999)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov

Tissue Procurement: CLONETECH Laboratories, Inc.
CDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

FEATURES
source
1. 999
Location/Qualifiers
Plate: L10M1066 row: h column: 10
High quality sequence stop: 562.

BASE COUNT 273 a 269 c 241 g 216 t
ORIGIN
1. 999
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4248249"
/clone_lib="NIH_MGC_83"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: prostate; Vector: pDNR-LIB (Clontech); Site_1: SfII (ggcgctcgcc); Site_2: SfII (ggccatagggc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGGCGGCGACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.4 kb (range 0.5-4.0 kb). 14/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

Query Match 7.8%; Score 89; DB 10; Length 999;
Best Local Similarity 100.0%; Pred. No. 3.le-32;
Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1053 agacagcgcgcgtccacatccagctcttggctcctcgctgctgtctctttccac 1112
|||||
Db 2 AGACAGGTCGCTCCACATCCAGGCTTTGGCTCCTCGCTGCTCTTTCCAC 61
|||||

QY 1113 gcatttccagataactgtgactccagg 1141
|||||
Db 62 GCATTTTCCAGGATAACTGTGACTCCAGG 90
|||||

RESULT 5
BG723255
LOCUS BG723255 714 bp mRNA linear EST 08-MAY-2001
DEFINITION 602690875F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4823059 5', mRNA sequence.
ACCESSION BG723255
VERSION BG723255.1 GI:14002442
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 114)
REFERENCE NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.

CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

FEATURES
source
1. 714
Location/Qualifiers
Plate: LLAM10732 row: f column: 20
High quality sequence stop: 711.

BASE COUNT 211 a 167 c 147 g 187 t
ORIGIN
1. 714
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4823059"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/note="Organ: testis; Vector: pBluescriptR (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag zdf62c07 r1 Soares fetal heart_NbHH19W Homo sapiens cDNA clone IMAGE:345328 5' similar to gb:U04313 MASPIN PRECURSOR (HUMAN)); size-selected for average insert size 2.2 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."

Query Match 7.3%; Score 83; DB 10; Length 714;
Best Local Similarity 100.0%; Pred. No. 2.6e-29;
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1059 gtcgcctccacatccagcttctgtcctcgctgctgtctctttccacgcat 1118
|||||
Db 6 GTGCGCTCCATCCAGGCTTTGGCTCCTCGCTGCTCTTTCCAGCATTT 65
|||||

QY 1119 tccaggataactgtgactccagg 1141
|||||
Db 66 TCCAGGATAACTGTGACTCCAGG 88
|||||

RESULT 6
W74001
LOCUS W74001 442 bp mRNA linear EST 20-JUN-1996
DEFINITION w62c07 r1 Soares fetal heart_NbHH19W Homo sapiens cDNA clone IMAGE:345328 5' similar to gb:U04313 MASPIN PRECURSOR (HUMAN);, mRNA sequence.
ACCESSION W74001
VERSION W74001.1 GI:1384551
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 442)
REFERENCE Hillier, L., Clark, N., Dubuque, T., Ellington, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Willson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: mob.REGA+ET
High quality sequence stop: 182.
Location/Qualifiers

Query Match 4.8%; Score 55; DB 12; Length 648;
 Best Local Similarity 100.0%; Pred. No. 1.1e-15;
 Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 327 tttttgtagagatgggtccaccatgttgcagagttggtctcaaacctcctggg 381
 |||||
 Db 562 TTTTGTAGAGATGGGCTCACCATTGTGGCCAGGTTGGTCAAACTCCTGGG 508

RESULT 9
 AL602554
 DEFINITION DKFZp686F1514_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
 DKFZp686F1514 5', mRNA sequence.
 ACCESSION AL602554
 VERSION AL602554.1 GI:15166060
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 625)
 Wambutt,R., Heubner,D., Mewes,W., Well,B. and Wiemann,S.
 EST (Wambutt,R., Heubner,D., Mewes,H.W., Well,B. and Wiemann,S.)
 Unpublished (1999)
 Contact: Wambutt R
 MIPS

Am Klopferspitz 18a D-82152 Martinsried, Germany
 This is the 5' sequence of the clone insert
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
 sequenced by AGOWA (Berlin/Germany) within the CDNA sequencing
 consortium of the German Genome Project.
 No sl sequence available.
 This clone (DKFZp686F1514) is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES
 source
 Location/Qualifiers
 1. .625
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="DKFZp686F1514"
 /clone_lib="686 (synonym: hlcc3)"
 /tissue_type="human skeletal muscle"
 /dev_stage="adult"
 /lab_host="DH108"
 /note="Vector: pTriplex2; Site_1: SfiIA; Site_2: SfiIB;
 cDNA-collection"
 BASE COUNT 180 a 149 c 122 g 174 t
 ORIGIN

Query Match 4.7%; Score 54; DB 9; Length 625;
 Best Local Similarity 100.0%; Pred. No. 3.5e-15;
 Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1088 cctcgctgcctgtctcttccagcatttccaggataactgtgactccagg 1141
 |||||
 Db 32 CCTCGCTGCCTGTCCTTTCCACGCATTTCCAGGATAACTGTGACTCCAGG 85

RESULT 10
 AA382130
 LOCUS
 DEFINITION EST95313 Activated T-cells II Homo sapiens cDNA 5' end similar to
 EST containing Alu repeat, mRNA sequence.
 ACCESSION AA382130
 VERSION AA382130.1 GI:2034469
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE AUTHORS

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 288)
 Adams M.D., Kerlavage, A.R., Fleischmann, R.D., Fulder, R.A., Bult
 C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White
 O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Wai, C., Clayton, R.A.,
 Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald,
 L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, N.S., Glodek, A.,
 Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S. Jr., Kelley, J.M.,
 Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M.,
 Moreno-Palancas, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M.,
 Phillips, C.A., Ryder, S.E., Scott, J.L., Saudek, D.M., Shirley, R.,
 Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y.,
 Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,
 Dinke, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A., He, W.W.,
 Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L.,
 Kunsch, C., Hung, J., Li, H., Meissner, P.S., Olsen, H., Raymond, L.,
 Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillon, P.J., Fannon
 M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and
 Venter, J.C.

TITLE
 Initial assessment of human gene diversity and expression patterns
 based upon 83 million nucleotides of cDNA sequence

JOURNAL
 Nature 377 (6547 Suppl), 3-174 (1995)

MEDLINE
 96026280

COMMENT
 Other ESTs: EST95312

Contact: Kerlavage, AR

Bioinformatics

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850 USA

Tel: 301899056

Fax: 301899423

Email: arkerlavet@tigr.org

For clone availability, additional sequence and expression

information related to this EST, please check the TIGR Human Gene

Index (<http://www.tigr.org/tdb/hgi/hgi.html>)

Seq primer: M13 Reverse

Location/Qualifiers

1. .288

/organism="Homo sapiens"

/db_xref="ATCC (Inhost):186511"

/db_xref="taxon:9606"

/clone_lib="Activated T-cells II"

/cell_type="T-lymphocyte"

/dev_stage="adult"

/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
 XhoI"

BASE COUNT 79 a 64 c 44 g 100 t 1 others

ORIGIN

Query Match 4.1%; Score 47; DB 9; Length 288;
 Best Local Similarity 100.0%; Pred. No. 8.8e-12;
 Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 346 tcaccatgttgcaggttggtctcaaacctcctggctcaagcaatc 392

|||||

Db 15 TCACCATGTGCCAGTGGTCTCAAACTCCTGGGCTCAAGCAATC 61

RESULT 11
 AA565270/c
 LOCUS
 DEFINITION nk53b11.s1 NCI-CGAP_Pr7 Homo sapiens cDNA clone IMAGE:1017213
 similar to contains Alu repetitive element;; mRNA sequence.

ACCESSION AA565270

VERSION AA565270.1 GI:2336909

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 382)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

TITLE

JOURNAL
COMMENT

Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: David G. Bostwick, M.D., Rodrigo F. Chuaqui,
M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 533 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 341.
Location/Qualifiers

source
1. .382
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="1017213"
/clone_lib="NCI_CGAP_Pr7"
/sex="male"
/tissue_type="low-grade prostatic neoplasia"
/lab_host="DH108"
/note="Organ: prostate; Vector: pAMP10; mRNA made from
prostate intraepithelial neoplasia (high-grade), cDNA
made by oligo-dT priming. Non-directionally cloned.
Size-selected on agarose gel, average insert size 600 bp."

BASE COUNT 112 a 77 c 92 g 101 t

ORIGIN

Query Match 3.9%; Score 45; DB 9; Length 382;
Best Local Similarity 100.0%; Pred. No. 8.3e-11;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 349 ccattgtccagggttgctcaaacctcctgggtcgaagcaatcc 393
|||||
Db 222 CCATGTTGCCAGGTGCTCAAACTCTCTGGGCTCAAGCAATCC 178.

RESULT 12
AA713761
LOCUS
DEFINITION
AA713761 395 bp mRNA linear EST 18-FEB-1998
nv82e02 s1 NCI_CGAP_Pr4 Homo sapiens cDNA clone IMAGE:1236314
similar to contains Alu repetitive element;; mRNA sequence.

ACCESSION
AA713761
VERSION
AA713761.1 GI:2726035
KEYWORDS
EST.
SOURCE
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 395)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Ilan Kirsch, M.D., Kristina A. Cole, M.D.,
Ph.D. student, Rodrigo F. Chuaqui, M.D., Michael R. Emmert-Buck,
M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 458 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham.
Location/Qualifiers

FEATURES
source

source
1. .395
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="1236314"
/clone_lib="NCI_CGAP_Pr4"
/sex="female"
/tissue_type="normal ductal tissue"
/lab_host="DH108"
/note="Organ: breast; Vector: pAMP10; mRNA made from
normal breast ductal tissue, cDNA made by oligo-dT
priming. Non-directionally cloned. Size-selected on
agarose gel, average insert size 600 bp. Reference:
Krizman et al. (1996) Cancer Research 56:5380-5383."
BASE COUNT 100 a 99 c 88 g 108 t

ORIGIN

Query Match 3.9%; Score 45; DB 9; Length 395;
Best Local Similarity 100.0%; Pred. No. 8.4e-11;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 349 ccattgtccagggttgctcaaacctcctgggtcgaagcaatcc 393
|||||
Db 206 CCATGTTGCCAGGTGCTCAAACTCTCTGGGCTCAAGCAATCC 250

RESULT 13
AA550845
LOCUS
DEFINITION
nj34e10.s1 NCI_CGAP_AAL Homo sapiens cDNA clone IMAGE:994410 3'
similar to contains Alu repetitive element;; mRNA sequence.

ACCESSION
AA550845
VERSION
AA550845.1 GI:2321097
KEYWORDS
EST.
SOURCE
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 484)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Stratagene, Inc., David B. Krizman,
Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 434.
Location/Qualifiers

FEATURES
source
1. .484
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="994410"
/clone_lib="NCI_CGAP_AAL"
/tissue_type="adrenal adenoma"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: adrenal gland; Vector: Bluescript SK-;
Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally.
Primer: Oligo dT. Two pooled bulk adrenal adenomas. 5'
adaptor sequence: 5' GAATTCGGCAGG 3' 3' adaptor
sequence: 5' CTCGAGTGTGTTTTTTTTTTT 3' Average insert
size: 1.6 kb."
BASE COUNT 130 a 104 c 99 g 151 t

ORIGIN

Search completed: September 24, 2002, 16:11:34
Job time: 8704 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: September 24, 2002, 10:15:05 ; Search time 3102.12 seconds
(without alignments)
7697.052 Million cell updates/sec

Title: US-09-617-174B-1

Perfect score: 1141

Sequence: 1 agataagcacagcagagaag.....aggataactgtgactccagg 1141

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*

1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description

1	1129.8	99.0	170269	9	AC090307	AC090307 Homo sapi
c	2	1129.8	99.0	173131	2	AC036176 Homo sapi
c	3	1129.8	99.0	173836	2	AP001404 Homo sapi
c	4	745.4	65.3	12426	6	AX251236 Sequence
c	5	723.4	63.4	12426	6	AX251235 Sequence
c	6	612	53.6	214212	2	AC015793 Homo sapi
c	7	437.6	38.4	214212	2	AC015793 Homo sapi
c	8	203.4	17.8	44255	9	AL356309 Human DNA
c	9	199.4	17.5	84577	9	AC079927 Homo sapi
c	10	195	17.1	2343	9	AK023951 Homo sapi
c	11	195	17.1	71872	9	AL133390 Human DNA
c	12	195	17.1	147259	2	AC019248 Homo sapi
c	13	195	17.1	184703	9	AL365215 Human DNA
c	14	194	17.0	164126	2	AC090104 Homo sapi
c	15	194	17.0	169669	2	AC041005 Homo sapi
c	16	194	17.0	190689	2	AC091185 Homo sapi
c	17	193.4	17.0	158806	2	AL591132 Homo sapi
c	18	193.4	17.0	168834	2	AL355583 Homo sapi
c	19	193.4	17.0	200882	2	AL591181 Homo sapi
c	20	193.4	17.0	214180	2	AL360082 Homo sapi
c	21	193.4	17.0	238379	2	AL590986 Homo sapi
c	22	193.4	17.0	300000	9	AL645813 Human DNA
c	23	192.8	16.9	161296	2	AC012308 Homo sapi
c	24	192.6	16.9	196533	9	AL096870 Human chr
c	25	191.6	16.8	148555	2	AP004711 Homo sapi
c	26	191.6	16.8	155614	2	AC012620 Homo sapi
c	27	191.6	16.8	160887	2	AC021498 Homo sapi
c	28	191.6	16.8	225037	2	AC091825 Homo sapi
c	29	191.4	16.8	162554	9	AC079354 Homo sapi
c	30	190.8	16.7	102837	9	HS2202 Human DNA s
c	31	190.8	16.7	183245	2	AC099397 Homo sapi
c	32	190.6	16.7	199179	2	AC024619 Homo sapi
c	33	190.4	16.7	109902	9	HS046C2 Human DNA
c	34	190.4	16.7	178964	2	AC012468 Homo sapi
c	35	190.2	16.7	150477	2	AC024441 Homo sapi
c	36	190.2	16.7	159265	2	AL157942 Homo sapi
c	37	190.2	16.7	168759	2	AC068307 Homo sapi
c	38	190.2	16.7	203918	2	AC084871 Homo sapi
c	39	190	16.7	110000	2	AL357131_0 Homo sapi
c	40	190	16.7	157044	2	AC079849 Homo sapi
c	41	190	16.7	180533	2	AL589943 Homo sapi
c	42	190	16.7	187052	2	AC078926 Homo sapi
c	43	190	16.7	188923	9	AC020612 Homo sapi
c	44	190	16.7	215867	2	AC084037 Homo sapi
c	45	190	16.7	257632	2	AC009246 Homo sapi

ALIGNMENTS

RESULT 1
AC090307 AC090307 170269 bp DNA linear PRI 30-DEC-2001
LOCUS Homo sapiens chromosome 18, clone RP11-851B10, complete sequence.
DEFINITION AC090307
ACCESSION AC090307
VERSION AC090307.7 GI:16041402
KEYWORDS HTG
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 170269)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 18, clone RP11-851B10
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 170269)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Glnde,S., Graham,L., Grand-pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,

Jones, C., Karatas, A., LaRocque, K., Lamazares, R., Landers, T.,
Lehocsky, J., Levine, R., Liu, G., MacLean, C., Macdonald, P.,
Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K.,
McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mienga, V.,
Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R.,
Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M.,
Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P.,
Sounguez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A.,
Willson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A., and Zody, M.

TITLE
JOURNAL
REFERENCE

Direct Submission
Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 170269)

3 (bases 1 to 170269)

Barren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouckhalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Gao, S., Gelfand, M., Graham, L., Grand-Pierre, N., Ginde, S., Gord, S., Goyette, M., Horton, L., Hulme, W., Iliev, T., Johnson, R., Jones, C., Keaford, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczeky, J., Levine, R., Liu, G., MacLean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McSwan, P., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Milnga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Riback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, J., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triglio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE	Direct Submission
JOURNAL	Submitted (30-DEC-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT	On Oct 11, 2001 this sequence version replaced gi:15290867.

COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu

FEATURES	Location/Qualifiers
source	1. .170269
repeat_region	/organism="Homo sapiens"
repeat_region	/db_xref="taxon:9606"
repeat_region	/chromosome="18"
repeat_region	/map="18"
repeat_region	/clone="RP11-851B10"
repeat_region	/clone_lib="RPC1-11 Human"
repeat_region	complement(2..177)
repeat_region	/rpt_family="ILMCS"
repeat_region	complement(189..489)
repeat_region	/rpt_family="AluXs"
repeat_region	complement(708..803)
repeat_region	/rpt_family="MIR"
repeat_region	complement(1616..1904)
repeat_region	/rpt_family="AluJo"
repeat_region	complement(2815..3120)

repeat_region	/rpt_family="AluX"	3218..3365
repeat_region	/rpt_family="MIR"	3565..3590
repeat_region	/rpt_family="(TG)n"	complement(3782..4028)
repeat_region	/rpt_family="MIR"	4056..4177
repeat_region	/rpt_family="MERSA"	4601..5620
repeat_region	/rpt_family="HSMAR2"	5621..5643
repeat_region	/rpt_family="(CAA)n"	5644..5866
repeat_region	/rpt_family="HSMAR2"	6184..6475
repeat_region	/rpt_family="LJMA2"	6677..6837
repeat_region	/rpt_family="(TA)n"	6852..7018
repeat_region	/rpt_family="(TA)n"	6994..7122
repeat_region	/rpt_family="(CATATA)n"	7146..7332
repeat_region	/rpt_family="(CATATA)n"	7371..7526
repeat_region	/rpt_family="(TA)n"	complement(7527..7835)
repeat_region	/rpt_family="AluX"	7936..8080
repeat_region	/rpt_family="MIR"	8697..8927
repeat_region	/rpt_family="AluX"	complement(9036..9288)
repeat_region	/rpt_family="AluSq"	9873..9934
repeat_region	/rpt_family="(GA)n"	10340..10367
repeat_region	/rpt_family="AT-rich"	10368..10686
repeat_region	/rpt_family="AluY"	10687..10711
repeat_region	/rpt_family="AT-rich"	complement(10870..11043)
repeat_region	/rpt_family="L2"	11348..11375
repeat_region	/rpt_family="(T)n"	12359..12449
repeat_region	/rpt_family="L3"	complement(12634..12890)
repeat_region	/rpt_family="MIR"	complement(13436..13725)
repeat_region	/rpt_family="AluJb"	14244..14297
repeat_region	/rpt_family="(CGGG)n"	complement(15188..15350)
repeat_region	/rpt_family="MIR3"	16056..16114
repeat_region	/rpt_family="MIR3"	complement(16129..16429)
repeat_region	/rpt_family="AluX"	16532..16646
repeat_region	/rpt_family="MER94"	16855..17145
repeat_region	/rpt_family="L2"	17377..17572
repeat_region	/rpt_family="MIR3"	complement(20036..20330)
repeat_region	/rpt_family="Charliel"	complement(20333..20486)
repeat_region	/rpt_family="Charliel"	20625..20691
repeat_region	/rpt_family="L2"	

```

repeat_region complement(20829..21075)
repeat_region /rpt_family="MLT1C"
repeat_region complement(21104..21336)
repeat_region /rpt_family="MLT1C"
repeat_region complement(22112..22391)
repeat_region /rpt_family="AluSx"
repeat_region 22506..22581
repeat_region /rpt_family="Tigger5"
repeat_region complement(23483..23664)
repeat_region /rpt_family="MIR"
repeat_region 23979..24190
repeat_region /rpt_family="MER20"
repeat_region complement(25404..25455)
repeat_region /rpt_family="MER5B"
repeat_region complement(25769..25853)
repeat_region /rpt_family="AluS"
unsure 25964..25998
/Note="Single clone coverage"
unsure 25999..26004
/Note="<30 qual single clone coverage"
unsure 26005..26009
/Note="Single clone coverage"
repeat_region complement(26204..26297)
repeat_region /rpt_family="MERSA"
repeat_region 27103..27222
repeat_region /rpt_family="MIR3"
repeat_region 27281..27573
repeat_region /rpt_family="AluY"
repeat_region 28033..28160
repeat_region /rpt_family="MIR"
repeat_region complement(28205..28490)
repeat_region /rpt_family="L1PB2"
repeat_region 28491..28801

Query Match 99.0%; Score 1129.8; DB 9; Length 170269;
Best Local Similarity 99.4%; Pred. No. 8e-309;
Matches 1134; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 agataagcacagagagaacacagctccgtttcaggtctcttctgaggtattcg 60
|||||
Db 13276 AGATAAGCACAGAGAGAACACAGCTCCGTTTCAGGTCTCTTCTCAGACTGATTCG 13335

QY 61 gctgaaggagtaggtcccccacaaatgaagaagctgtgtggaagacagagagacaagaac 120
|||||
Db 13336 GCTGAAGGAGTAGGTCCCGCAATGAAGAAGCTGTGGAGACAGAGGACAGAAGAAC 13395

QY 121 aggtctccacagagatttcacagacagctgctactctcttcttcttcttctt 180
|||||
Db 13396 AGGCTCCAGAACAGATTTTCAGAGCAGAGCTGCGTACTCTCTTTCTTTTCTTTT 13455

QY 181 gctctgacccagagctgaagtacagtgttagctcagcgtcactgcaagcttgacctc 240
|||||
Db 13456 GCTCTGTACCCAGCGTGAAGTACAGTGTGTAGCTCAGCGCTCAGTGCAGCTTTGACCTC 13515

QY 241 ccaggctcaagtgtactctctctcagcttctccaaagtaactgggaccacagggcatgcac 300
|||||
Db 13516 CCAGGCTCAAGTGATCCTCTGCTCAGCTTTCCAAAGTAACCTGGGACCACAGCATGCAT 13575

QY 301 caccacgttagctattgttttacatttttttagagatggggtctcaccatgttgcacca 360
|||||
Db 13576 CACCACATAGGCTATGTGTTTACATTTTGTAGAGATGGGGTCTCACCATGTTGCCCA 13635

QY 361 ggtggtctcaactctggctcagcaatccgctcagctcaactcccccaaatgctgg 420
|||||
Db 13636 GGTGTGCTCAAACTCTGGGCTCAAGCAATCGCTCAGCTCAACCTCCCAAAATGCTGG 13695

QY 421 gattacagcgtagcaccggccagggctgagtaactcctaatacagagattttaaaaa 480
|||||
Db 13696 GATTACAGCGTGAGCCACCGCCAGGCTTGAGTAATCCTAATCACAGGATTTTAAAAA 13755

QY 481 gaaactctgcgcacccattaaacaatatctctaccaatcttgtagtaaatatttg 540
|||||
Db 13756 GAAACTTCTGCGCCACCCATTAAACAATATCTCTACCAATTTGTTAGTAAATATTTTG 13815

```

```

QY 541 ctaatagtacctaatttttaggtaggcactgtgtttatacatatataccattcttcttt 600
|||||
Db 13816 CTAATAGTACCTAATTTTAGGTAGGCACGTGTTATATACATATATCCATTCTTCTTTT 13875

QY 601 ttgattgttcttctttaaattgggcagctacatctcttggcatcttagcagaatagctgc 660
|||||
Db 13876 TTGATTGTCTTTCTGTTAATGGGAGCTACCTCTCTTGGCATCTAGCAGAATAGAGCTGC 13935

QY 661 tcagttttacacaaaaaagaatggagatcagagatcttttggccacacagctgctgag 720
|||||
Db 13936 TGCAGTTTACACAAAAGAATGGAGTACAGTACTTTTGTGCCAACACGTGCTGAG 13995

QY 721 aaattgtagtgttactatcatcacatttctttatttatttcgaataatttcaacctcc 780
|||||
Db 13996 AAATTGTAGTCTTACTATCATCACATTACTTTTATTTTCATCGAATATTTTCACTTCC 14055

QY 781 ggtcctgogtgggcccagagattgcgtacgcagtctgtacgtatgcatgtaactcac 840
|||||
Db 14056 GGTCTCGTGGCGGCGAGAGATTGCCGTACGCATGCTGTGATGATGATGATGATGATGATGAT 14115

QY 841 agcccttctctccggaacatgttgaggccttttggagccttttggagcctgtgcagacaacac 900
|||||
Db 14116 AGCCCTTCTCTCCCGGACATGTTGGAGGCTTTTGGAAAGCTGTGCAGACAAACAGTAAC 14175

QY 901 tcagctgaatcatctcttcaattgtggacaagctgccaagagcgttgcagtaggagag 960
|||||
Db 14176 TCAGCTGAATCATTTCTTCAATTTGTGCAGAGCTGCCAAGAGGCTTTCAGTAGGAGAG 14235

QY 961 agtgcgcgcagggcgggcgggcggtgtgagctgggctgggctgggctgggctgggctgggct 1020
|||||
Db 14236 AGTGC CGCGAGGCGGGCGGGCGGGCGGTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGG 14295

QY 1021 gctgccaggtgagcacacgcgtcttctccagacacgctgcctccacatccaggtct 1080
|||||
Db 14296 GCTGCCAGGTGAGCACACCGCTCTCTGCCAGACACGCTGCCCTCCACATCCAGGTCT 14355

QY 1081 ttgtgctctgctgtcctgttcttcttccacgcattttccaggataaactgtaactccag 1140
|||||
Db 14356 TTGTGCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 14415

QY 1141 g 1141
Db 14416 G 14416

RESULT 2
AC036176/c
LOCUS
DEFINITION Homo sapiens chromosome 18 clone RP11-635N19 map 18, *** SEQUENCING
IN PROGRESS ***, 2 ordered pieces.
AC036176
AC036176.7 GI:17647019
VERSION
KEYWORDS HTG: HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 173131)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abramson,H., Allen,N.,
TITLE Homo sapiens chromosome 18, clone RP11-635N19
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 173131)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abramson,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bedalov,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campoliano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Harford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,

```


Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 173836)
Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,
Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
Homo sapiens 173,836 genomic DNA of 18q22
Published Only in Database (2000) In press
2 (bases 1 to 173836)
Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,
Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
Direct Submission
Submitted (14-MAR-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
(E-mail: hattori@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
Tel: 81-42-778-9923, Fax: 81-42-778-9924)
On Jul 14, 2000 this sequence version replaced gi:8117315.
----- Genome Center -----

Center: RIKEN Genomic Sciences Center (GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information -----
Center project name: HumDraft18
Center clone name: RP11-851B10
----- Summary Statistics -----
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 171265 bases at least Q40
Consensus quality: 172100 bases at least Q30
Consensus quality: 172469 bases at least Q20
Insert size: 172836; sum-of-contigs
Quality coverage: 10.50x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of
11 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved

1 68896 contig of 68896 bp in length
68997 95899 contig of 26903 bp in length
96000 119268 contig of 23269 bp in length
119369 141739 contig of 23271 bp in length
141840 152655 contig of 10816 bp in length
152756 159567 contig of 6711 bp in length
159567 163465 contig of 3899 bp in length
163566 166433 contig of 2868 bp in length
166534 170112 contig of 3579 bp in length
170213 172095 contig of 1883 bp in length
172196 173836 contig of 1641 bp in length.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 68896: contig of 68896 bp in length
68997 68996: gap of 100 bp
95899 95899: contig of 26903 bp in length
96000 95999: gap of 100 bp
119268 119268: contig of 23269 bp in length
119369 119368: gap of 100 bp
141739 141739: contig of 23271 bp in length
141740 141839: gap of 100 bp
141840 152655: contig of 10816 bp in length
152656 152755: gap of 100 bp
152756 159466: contig of 6711 bp in length
159467 159566: gap of 100 bp

* 159567 163465: contig of 3899 bp in length
* 163466 163565: gap of 100 bp
* 163566 166433: contig of 2868 bp in length
* 166434 166533: gap of 100 bp
* 166534 170112: contig of 3579 bp in length
* 170113 170212: gap of 100 bp
* 170213 172095: contig of 1883 bp in length
* 172096 172195: gap of 100 bp
* 172196 173836: contig of 1641 bp in length.
FEATURES
Location/Qualifiers
1..173836
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18q22"
/clone="RP11-851B10"
1..68896
/note="assembly_fragment"
68997..95899
/note="assembly_fragment clone_end:T7 vector_side:right"
96000..119268
/note="assembly_fragment"
119369..141739
/note="assembly_fragment"
141840..152655
/note="assembly_fragment"
152756..159466
/note="assembly_fragment"
159567..163465
/note="assembly_fragment"
163566..166433
/note="assembly_fragment clone_end:SP6 vector_side:left"
166534..170112
/note="assembly_fragment"
170213..172095
/note="assembly_fragment"
172196..173836
/note="assembly_fragment"
BASE COUNT 53047 a 34410 c 33807 g 51570 t 1002 others
ORIGIN

Query Match 99.0%; Score 1129.8; DB 2; Length 173836;
Best Local Similarity 99.4%; Pred. No. 8.1e-309;
Matches 1134; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 1 agataagcacagcagagaagcaccagctccgtttcaggtcctcttctcaggtgattcg 60
|||||
Db 150724 AGATAAGCACAGCAGAGAAGCAACCAAGCTCCGTTTCAGGTCTCTTCTGAGACTGATTCG 150665
QY 61 gctggaaggagtagtcccacccaaatgaagaagctgtggagagacagagacaagaac 120
|||||
Db 150664 GCTGGGAAGGAGTAGTCCCGCCAAATGAAGAAGCTGTGGGAAGACAGAGGACAAGAAC 150605
QY 121 aggtccacagaagagatttcagagcagagctgcgtactcctttctttttttttttt 180
|||||
Db 150604 AGGCTCCACGAAGAGATTTTCAGAGCAGAGCTGCGTACTCTCTTTTCTTTTCTTTT 150545
QY 181 gctctgtaccacaggtggaagtacagtggttagctcaagctcagcttgccttc 240
|||||
Db 150544 GCTCTGTACCCAGGCTGAAGTACAGTGGTTAGCTCAGCGCTCACTGACGCTTGACCTC 150485
QY 241 ccaggtcgaagtagctctcgtctcagctttccagtaactgggaccacagcagcatgc 300
|||||
Db 150484 CCAGGCTCAAGTAGATCTCTCGTCTCAGCTTTCCAAAGTAAGTGGGACCAACAGCATGCAT 150425
QY 301 caccacgttaggcattgttttacatttttttagagatggggtctccacattgtccca 360
|||||
Db 150424 CACCACACTAGGCTATTGTTTACATTTTGTAGAGATGGGGCTCTCACCATTGTCCCA 150365
QY 361 ggttggtctcaaaactcctgggctcaagcaatccgcgtcaacgtccccaaatgctgg 420
|||||
Db 150364 GGTGTGCTCTCAAACTCCTGGGCTCAAGCAATCGGCTCAGCTCAACCTCCCAAAATGCTGG 150305

/db_xref="taxon:32630"	
/note="chemically treated genomic DNA (Homo sapiens)"	
BASE COUNT	3791 a 137 c 2517 g 5981 t
ORIGIN	
Query Match 65.3%; Score 745.4; DB 6; Length 12426;	
Best Local Similarity 78.4%; Pred. No. 3.7e-200;	
Matches 893; Conservative 0; Mismatches 246; Indels 0; Gaps 0;	
QY 1	agataagcaagcagagagaagaacacagctcgcgtttccaggtctcttccctcagagctgattcg 60
DB 8506	AAATAAACACACAAAAAACAACCACTCCGTTTCAATCTCTTCCCTAAACATAATTCG 8447
QY 61	gctgaaagggagtaggtccaccacaaatgaagaagctgtgggaagacagagagacaagaac 120
DB 8446	ACTAAAAAATAAATCCGCCCAATTAATAAACTATAAAAAACAACAAAAACAACAAAC 8387
QY 121	aggctccacgaagagatttcagagcagagctgctactcctcttcttcttcttcttcttctt 180
DB 8386	AAATCCACGAAAAAATTTCAAAACAAACTACGTACTCTCTTTTCTTTTATTTCTTTT 8327
QY 181	gctctgcaaccaggtgaagtacagtggttagctcagcgtcactcagcttgaactc 240
DB 8326	ACTCTATCACCACAACTAAATACAAATTAATTAACACGACTCAGTCACTTAACCTC 8267
QY 241	ccaggtcgaagtgcctcgtcagcttccaaagtaactgggacacagggcatgcat 300
DB 8266	CCAACTCAATAATCTCTCGTCTCAACTTTTCCAAATAACTTAAACACCAACATACAT 8207
QY 301	caccacgtcagctattgtttacatttttttagagatggggtctcaccatgttgcaca 360
DB 8206	CACACACTAAACTATATTTTACATTTTTTATAAAATAAATAATCTACCATATTAACCA 8147
QY 361	ggtgtgtcacaactcctgggtcgaagcaatccgcgtccacgtccaaactccccaaatgtc 420
DB 8146	AATTAATCTCAAACTCTTAACCTCAAAACAATCGCTCAGTCAACCTCCCAAACTAA 8087
QY 421	gattacagcgtgagccacccggcggtgagtaactcactcaacagagatttataaaa 480
DB 8086	AATTACAAACGTAACACCGCGCAACCTAAATAATCTTAATCAACAAATTTTAAAAA 8027
QY 481	gaactctcgtccaccacatatacaatctcctaccacatttggtagtaaatatttg 540
DB 8026	AAACCTTCTTACGCCACCCATTAACAATAATCTCTTACCAATTTAATAATAATTTTA 7967
QY 541	ctaagtactcaatttttagtgaggaactgtgtttatacatatccactcctcttctt 600
DB 7966	CTAATAATACCTAAATTTTAAATAAACACTATATTTATACATATATATCTCTCTTT 7907
QY 601	ttgattgtcttctgtttaatggcgagctacctctcttggtcactcagcagaatgagctgc 660
DB 7906	TTAATATCTTTCTTATTTTAAATAAACACTACCTCTCTTAACATCTAACAAAAATAACTAC 7847
QY 661	tgcaagttcacaaaaagaatggagatcagagctacttttggccaccacagtggtcag 720
DB 7846	TACAAATTACAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 7787
QY 721	aaatttgtagtgtaactatcatcacatttacttttttttttttttttttttttttttt 780
DB 7786	AAATTTATAATATATCTATCATCATCATCATCTTTTATTTTATCTCGAATATTTTCCCTTC 7727
QY 781	ggctcgtggtggcgagagattgcgtgacgcatgtctgtacgtatgcgtgtaactcac 840
DB 7726	GATCTAGTAAACCGAAAAAATTAACCGTAGCATATCTATAGTATACATAATAACTCAC 7667
QY 841	agcccttctcgtcccgacaatgttggaggccttttggagagctgtgcagacaacagcaact 900
DB 7666	AACCCCTTCTTACCGCAACATATTAATAAACTTTTAAAAAACTATACAAAAACAATAAAT 7607
QY 901	tcagcctgaatcatcttcttcaattgtggacaagctcccaagggctgtgagtaggagag 960
DB 7606	TCAACCTAAATCATTTCTTTCAATTAATAAACTACCAAACTACCAAACTAAATAAAAAA 7547

Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
 Lehoczy, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
 McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
 Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
 Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.

Direct Submission
 Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

On Mar 3, 2000 this sequence version replaced gi:5446894.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WBIR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L1590

Center clone name: L117_D_15

NOTE: This record contains 240 individual
 sequencing reads that have not been assembled into
 contigs. Runs of N are used to separate the reads
 and the order in which they appear is completely
 arbitrary. Low-pass sequence sampling is useful for
 identifying clones that may be gene-rich and allows
 overlap relationships among clones to be deduced.
 However, it should not be assumed that this clone
 will be sequenced to completion. In the event that
 the record is updated, the accession number will
 be preserved.

1 695: contig of 695 bp in length
 696 795: gap of 100 bp
 796 1513: contig of 718 bp in length
 1514 1613: gap of 100 bp
 1614 2337: contig of 724 bp in length
 2338 2437: gap of 100 bp
 2438 3161: contig of 724 bp in length
 3162 3261: gap of 100 bp
 3262 3966: contig of 705 bp in length
 3967 4066: gap of 100 bp
 4067 4772: contig of 706 bp in length
 4773 4872: gap of 100 bp
 4873 5564: contig of 692 bp in length
 5565 5664: gap of 100 bp
 5665 6397: contig of 733 bp in length
 6398 6497: gap of 100 bp
 6498 7202: contig of 705 bp in length
 7203 7302: gap of 100 bp
 7303 8028: contig of 726 bp in length
 8029 8128: gap of 100 bp
 8129 8803: contig of 675 bp in length
 8804 8903: gap of 100 bp
 8904 9613: contig of 710 bp in length
 9614 9713: gap of 100 bp
 9714 10409: contig of 696 bp in length
 10410 10509: gap of 100 bp
 10510 11237: contig of 728 bp in length
 11238 11337: gap of 100 bp
 11338 12043: contig of 706 bp in length
 12044 12143: gap of 100 bp
 12144 12867: contig of 724 bp in length
 12868 12967: gap of 100 bp
 12968 13660: contig of 693 bp in length
 13661 13760: gap of 100 bp
 13761 14489: contig of 729 bp in length
 14490 14589: gap of 100 bp
 14590 15314: contig of 725 bp in length

15315 15414: gap of 100 bp
 15415 16129: contig of 715 bp in length
 16130 16229: gap of 100 bp
 16230 16946: contig of 717 bp in length
 16947 17046: gap of 100 bp
 17047 17765: contig of 723 bp in length
 17770 17869: gap of 100 bp
 17870 18574: contig of 705 bp in length
 18575 18674: gap of 100 bp
 18675 19379: contig of 705 bp in length
 19380 19479: gap of 100 bp
 19480 20186: contig of 707 bp in length
 20187 20286: gap of 100 bp
 20287 20980: contig of 694 bp in length
 20981 21080: gap of 100 bp
 21081 21779: contig of 699 bp in length
 21780 21879: gap of 100 bp
 21880 22594: contig of 715 bp in length
 22595 22694: gap of 100 bp
 22695 23399: contig of 705 bp in length
 23400 23499: gap of 100 bp
 23500 24217: contig of 718 bp in length
 24218 24317: gap of 100 bp
 24318 25026: contig of 709 bp in length
 25027 25126: gap of 100 bp
 25127 25836: contig of 710 bp in length
 25837 25936: gap of 100 bp
 25937 26637: contig of 701 bp in length
 26638 26737: gap of 100 bp
 26738 27446: contig of 709 bp in length
 27447 27546: gap of 100 bp
 27547 28274: contig of 728 bp in length
 28275 28374: gap of 100 bp
 28375 29095: contig of 721 bp in length
 29096 29195: gap of 100 bp
 29196 29917: contig of 722 bp in length
 29918 30017: gap of 100 bp
 30018 30724: contig of 707 bp in length
 30725 30824: gap of 100 bp
 30825 31529: contig of 705 bp in length
 31530 31629: gap of 100 bp
 31630 32324: contig of 695 bp in length
 32325 32424: gap of 100 bp
 32425 33116: contig of 692 bp in length
 33117 33216: gap of 100 bp
 33217 33936: contig of 720 bp in length
 33937 34036: gap of 100 bp
 34037 34750: contig of 714 bp in length
 34751 34850: gap of 100 bp
 34851 35578: contig of 728 bp in length
 35579 35678: gap of 100 bp
 35679 36393: contig of 715 bp in length
 36394 36493: gap of 100 bp
 36494 37203: contig of 710 bp in length
 37204 37303: gap of 100 bp
 37304 37965: contig of 662 bp in length
 37966 38065: gap of 100 bp
 38066 38781: contig of 716 bp in length
 38782 38881: gap of 100 bp
 38882 39590: contig of 709 bp in length
 39591 39690: gap of 100 bp
 39691 40403: contig of 713 bp in length
 40404 40503: gap of 100 bp
 40504 41238: contig of 735 bp in length
 41239 41338: gap of 100 bp
 41339 42064: contig of 726 bp in length
 42065 42164: gap of 100 bp
 42165 42880: contig of 716 bp in length
 42881 42980: gap of 100 bp
 42981 43687: contig of 707 bp in length
 43688 43787: gap of 100 bp
 43788 44504: contig of 717 bp in length
 44505 44604: gap of 100 bp

* 44605	45307: contig of 703 bp in length
* 45308	45407: gap of 100 bp
* 45408	46106: contig of 699 bp in length
* 46107	46206: gap of 100 bp
* 46207	46809: contig of 703 bp in length
* 46910	47009: gap of 100 bp
* 47010	47696: contig of 687 bp in length
* 47697	47796: gap of 100 bp
* 47797	48531: contig of 735 bp in length
* 48532	48631: gap of 100 bp
* 48632	49341: contig of 710 bp in length
* 49342	49441: gap of 100 bp
* 49442	50162: contig of 721 bp in length
* 50163	50262: gap of 100 bp
* 50263	50990: contig of 728 bp in length
* 50991	51090: gap of 100 bp
* 51091	51803: contig of 713 bp in length
* 51804	51903: gap of 100 bp
* 51904	52610: contig of 707 bp in length
* 52611	52710: gap of 100 bp
* 52711	53412: contig of 702 bp in length
* 53413	53512: gap of 100 bp
* 53513	54253: contig of 741 bp in length
* 54254	54353: gap of 100 bp
* 54354	55084: contig of 731 bp in length
* 55085	55184: gap of 100 bp
* 55185	55972: contig of 788 bp in length
* 55973	56072: gap of 100 bp
* 56073	56855: contig of 783 bp in length
* 56856	56955: gap of 100 bp
* 56956	57708: contig of 753 bp in length
* 57709	57808: gap of 100 bp
* 57809	58579: contig of 771 bp in length
* 58580	58679: gap of 100 bp
Query Match	
Best Local Similarity 53.6%; Score 612; DB 2; Length 214212;	
Matches 633; Conservative 0; Mismatches 136; Indels 0; Gaps 0;	
QY 1	agataagcacagagagaacacagctccgttttcaggtctcttcctgagctgattcg 60
Db 168605	AGATAAGCACAGAGAGAAGCAACAGCTCTTTTCAGGTCTCTTCTCGAGACTGATTCG 168546
QY 61	gctggaaggagtaggtccaccacaaatgaagagctgtgggaagacagagagacaagaac 120
Db 168545	GCTGGAAGGAGTAGGTCCCGCCCAATGAGAAAGCTGTGGGAAGACAGAGGACAGAAC 168486
QY 121	aggctccagaagagatttcagagcagagctgcgtactccttttttttttttttttttt 180
Db 168485	AGGCTCCAGCAAGAGATTCAGAGCAGAGCTGCTACTCTTTTCTTTTGTCTTTT 168426
QY 181	gctctgaccagagctgaagtacagtggttagctcacgctcactgcagcttgacctc 240
Db 168425	GCTCTGTACCCAGGCTGAAGTACAGTGTGTAGCTTACGCTCACGGCTACATGACCTC 168366
QY 241	ccaggctcaagtatctctctcagctttccaaagtaactgggaaccacagcagtcgat 300
Db 168365	CCAGGCTCAAGTATCTCTCTCAGCTTTCCAAAGTAACTGGGACCACAGGCATGCAT 168306
QY 301	caccacgctaggctattgttttacattttttgtagagatggggtctcaccattgtgccca 360
Db 168305	CACCACACTAGGCTATTGTTTACATTTTGTAGAGATGGGCTCTCAGCATGTTGCCCA 168246
QY 361	ggttggtctcaaaactcctgggtccaagaatccgctcaagtcacaccccccacaaagtgg 420
Db 168245	GGTTGCTCTCAAACTCCTGGGCTCAAGCAATCGCTCACGTCAACCTCCCAATGCTGG 168186
QY 421	gattacagcggtgagccacggcgagtgagtaactcctaaccagagatttataaaa 480
Db 168185	GATTACAGCGGTGAGCCACCTCGCCAGGCTGAGTAACTCTATACAGAGATTTTAAAAA 168126
QY 481	gaaactctcgcgcacccattaaaacaatatctcctaccaatttggtagtaaatatttg 540

Db 168125	GAACCTTCCTGCGCCACCCATTAACAAATATCTCTACCAATTTGGTAGTAATATTTTG 168066
QY 541	ctaaatagctacaaatttttagtagtagcactgtgtttatatacatatcattctctttt 600
Db 168065	CYAAATAGTACCTAAATTTTATAGGTAGCACTGTGTTATACATATATCCATTCTCTTTT 168006
QY 601	ttgattgtctctctgttttaattggcgagctacctctctctgtgcattctagcagaatgagctgc 660
Db 168005	TTGATTGTCTCTTCTGTTTAATCAATCCACGACATTCGACCTTCAGCTGGTGNNNNNN 167946
QY 661	tgcaatttacacaaaagaatggagatcagatcattcttttgcaccacacgtctgtag 720
Db 167945	NN 167886
QY 721	aaattgttagtgttactatcatcacacattcttttatttatttcaata 769
Db 167885	NN 167837

RESULT 7
LOCUS AC015793 214212 bp DNA linear HTG 13-JUL-2000
DEFINITION Homo sapiens chromosome 18 clone RP11-1117D15 map 18, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC015793
VERSION AC015793.2 GI:7144769
KEYWORDS HTG: HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 214212)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 18, clone RP11-1117D15
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 214212)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collamore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
DIRECT SUBMISSION
Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 3, 2000 this sequence version replaced gi:6446894.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1590
Center clone name: 1117_D_15

* NOTE: This record contains 240 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.

* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

* 1
* 696 795: gap of 100 bp
* 796 1513: contig of 718 bp in length
* 1514 1613: gap of 100 bp
* 1614 2337: contig of 724 bp in length
* 2338 2437: gap of 100 bp
* 2438 3161: contig of 724 bp in length
* 3162 3261: gap of 100 bp
* 3262 3966: contig of 705 bp in length
* 3967 4066: gap of 100 bp
* 4067 4772: contig of 706 bp in length
* 4773 4872: gap of 100 bp
* 4873 5564: contig of 692 bp in length
* 5565 5664: gap of 100 bp
* 5665 6397: contig of 733 bp in length
* 6398 6497: gap of 100 bp
* 6498 7202: contig of 705 bp in length
* 7203 7302: gap of 100 bp
* 7303 8028: contig of 726 bp in length
* 8029 8128: gap of 100 bp
* 8129 8803: contig of 675 bp in length
* 8804 8903: gap of 100 bp
* 8904 9613: contig of 710 bp in length
* 9614 9713: gap of 100 bp
* 9714 10409: contig of 696 bp in length
* 10410 10509: gap of 100 bp
* 10510 11237: contig of 728 bp in length
* 11238 11337: gap of 100 bp
* 11338 12043: contig of 706 bp in length
* 12044 12143: gap of 100 bp
* 12144 12867: contig of 724 bp in length
* 12868 12967: gap of 100 bp
* 12968 13660: contig of 693 bp in length
* 13661 13760: gap of 100 bp
* 13761 14489: contig of 729 bp in length
* 14490 14589: gap of 100 bp
* 14590 15314: contig of 725 bp in length
* 15315 15414: gap of 100 bp
* 15415 16129: contig of 715 bp in length
* 16130 16229: gap of 100 bp
* 16230 16946: contig of 717 bp in length
* 16947 17046: gap of 100 bp
* 17047 17769: contig of 723 bp in length
* 17770 17869: gap of 100 bp
* 17870 18574: contig of 705 bp in length
* 18575 18674: gap of 100 bp
* 18675 19379: contig of 705 bp in length
* 19380 19479: gap of 100 bp
* 19480 20186: contig of 707 bp in length
* 20187 20286: gap of 100 bp
* 20287 20980: contig of 694 bp in length
* 20981 21080: gap of 100 bp
* 21081 21779: contig of 699 bp in length
* 21780 21879: gap of 100 bp
* 21880 22594: contig of 715 bp in length
* 22595 22694: gap of 100 bp
* 22695 23398: contig of 705 bp in length
* 23400 23499: gap of 100 bp
* 23500 24217: contig of 718 bp in length
* 24218 24317: gap of 100 bp
* 24318 25026: contig of 709 bp in length
* 25027 25126: gap of 100 bp
* 25127 25836: contig of 710 bp in length
* 25837 25936: gap of 100 bp
* 25937 26637: contig of 701 bp in length
* 26638 26737: gap of 100 bp
* 26738 27446: contig of 709 bp in length
* 27447 27546: gap of 100 bp
* 27547 28274: contig of 728 bp in length

* 28275 28374: gap of 100 bp
* 28375 29095: contig of 721 bp in length
* 29096 29195: gap of 100 bp
* 29196 29917: contig of 722 bp in length
* 29918 30017: gap of 100 bp
* 30018 30724: contig of 707 bp in length
* 30725 30824: gap of 100 bp
* 30825 31529: contig of 705 bp in length
* 31530 31629: gap of 100 bp
* 31630 32324: contig of 695 bp in length
* 32325 32424: gap of 100 bp
* 32425 33116: contig of 692 bp in length
* 33117 33216: gap of 100 bp
* 33217 33936: contig of 720 bp in length
* 33937 34036: gap of 100 bp
* 34037 34750: contig of 714 bp in length
* 34751 34850: gap of 100 bp
* 34851 35578: contig of 728 bp in length
* 35579 35678: gap of 100 bp
* 35679 36393: contig of 715 bp in length
* 36394 36493: gap of 100 bp
* 36494 37203: contig of 710 bp in length
* 37204 37303: gap of 100 bp
* 37304 37965: contig of 662 bp in length
* 37966 38065: gap of 100 bp
* 38066 38781: contig of 716 bp in length
* 38782 38881: gap of 100 bp
* 38882 39590: contig of 709 bp in length
* 39591 39690: gap of 100 bp
* 39691 40403: contig of 713 bp in length
* 40404 40503: gap of 100 bp
* 40504 41238: contig of 735 bp in length
* 41239 41338: gap of 100 bp
* 41339 42064: contig of 726 bp in length
* 42065 42164: gap of 100 bp
* 42165 42880: contig of 716 bp in length
* 42881 42980: gap of 100 bp
* 42981 43687: contig of 707 bp in length
* 43688 43787: gap of 100 bp
* 43788 44504: contig of 717 bp in length
* 44505 44604: gap of 100 bp
* 44605 45307: contig of 703 bp in length
* 45308 45407: gap of 100 bp
* 45408 46106: contig of 699 bp in length
* 46107 46206: gap of 100 bp
* 46207 46909: contig of 703 bp in length
* 46910 47009: gap of 100 bp
* 47010 47696: contig of 687 bp in length
* 47697 47796: gap of 100 bp
* 47797 48531: contig of 735 bp in length
* 48532 48631: gap of 100 bp
* 48632 49341: contig of 710 bp in length
* 49342 49441: gap of 100 bp
* 49442 50162: contig of 721 bp in length
* 50163 50262: gap of 100 bp
* 50263 50990: contig of 728 bp in length
* 50991 51090: gap of 100 bp
* 51091 51803: contig of 713 bp in length
* 51804 51903: gap of 100 bp
* 51904 52610: contig of 707 bp in length
* 52611 52710: gap of 100 bp
* 52711 53412: contig of 702 bp in length
* 53413 53512: gap of 100 bp
* 53513 54253: contig of 741 bp in length
* 54254 54353: gap of 100 bp
* 54354 55084: contig of 731 bp in length
* 55085 55184: gap of 100 bp
* 55185 55972: contig of 788 bp in length
* 55973 56072: gap of 100 bp
* 56073 56855: contig of 783 bp in length
* 56856 56955: gap of 100 bp
* 56956 57708: contig of 753 bp in length
* 57709 57808: gap of 100 bp

RESULT 9

AC079927 84577 bp DNA linear PRI 30-JAN-2002
 LOCUS Homo sapiens BAC clone RP11-702A23 from 4, complete sequence.
 DEFINITION AC079927
 ACCESSION AC079927.5 GI:18071472
 VERSION
 HTG.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 84577)
 Sulston, J.E. and Waterston, R.

Toward a complete human genome sequence
 Genome Res. 8 (11), 1097-1108 (1998)
 99063792

2 (bases 1 to 84577)
 Cedroni, M., Abbott, A. and Boyer, E.

The sequence of Homo sapiens BAC clone RP11-702A23
 Unpublished (2001)

3 (bases 1 to 84577)
 Waterston, R.H.

Direct Submission
 Submitted (17-SEP-2000) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

4 (bases 1 to 84577)
 Waterston, R.H.

Direct Submission
 Submitted (05-JAN-2002) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

5 (bases 1 to 84577)
 Waterston, R.

Direct Submission
 Submitted (30-JAN-2002) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Jan 5, 2002 this sequence version replaced gi:16924160.

----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc>

Contact: sapiens@watson.wustl.edu

----- Summary Statistics

----- Center project name: H_NH0702A23

NOTICE: This sequence may not represent the entire insert of this
 clone. It may be shorter because we only sequence overlapping
 clone sections once, or longer because we provide a small overlap
 between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate
 chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by sequence
 from more than one subclone; and the assembly was confirmed by
 restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D.
 McPherson, Department of Genetics, Washington University, St. Louis
 MO. For additional information about the map position of this
 sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPC1-11 human BAC library was made from the blood of one male
 donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E.,
 Tatenno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved
 approach for construction of bacterial artificial chromosome
 libraries. Genomics 51:1-8. The clone may be obtained either from

Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong
 and coworkers at the Roswell Park Cancer Institute
 (<http://bacpac.med.buffalo.edu>)
 VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-757K22, 2000 bp overlap;
 the clone sequenced to the right is RP11-19F13. Actual start of
 this clone is at base position 110812 of RP11-757K22; actual end is
 at base position 84577 of RP11-702A23.

Two transposons occurred in the growth of the clone that have been
 omitted from the finished sequence.

FEATURES	Source	Location/Qualifiers
repeat_region	1..84577	/organism="Homo sapiens"
repeat_region	256..528	/db_xref="taxon:9606"
repeat_region	697..723	/chromosome="4"
repeat_region	713..846	/map="4"
repeat_region	1407..1747	/clone="RP11-702A23"
repeat_region	1799..2262	/clone_lib="RPC1-11"
repeat_region	2265..2290	1..206
repeat_region	2266..2574	/rpt_family="Alu"
repeat_region	2599..2719	/rpt_family="MER1_type"
repeat_region	2891..2996	/rpt_family="L2"
repeat_region	3255..3607	/rpt_family="TTTn"
repeat_region	3620..3713	/rpt_family="MER2_type"
repeat_region	3637..3719	/rpt_family="MER53"
repeat_region	4055..4083	/rpt_family="AT_rich"
repeat_region	4056..4358	/rpt_family="Alu"
repeat_region	4698..4721	/rpt_family="AT_rich"
repeat_region	5095..5462	/rpt_family="MaLR"
repeat_region	6970..7010	/rpt_family="(TTC)n"
repeat_region	6985..7294	/rpt_family="Alu"
repeat_region	7703..7992	/rpt_family="Alu"
repeat_region	8526..8559	/rpt_family="(T)n"
repeat_region	8532..8840	/rpt_family="Alu"
repeat_region	9438..9458	/rpt_family="MIR"
repeat_region	9440..9738	/rpt_family="AT_rich"
repeat_region	9744..9886	/rpt_family="Alu"
repeat_region	10404..10507	/rpt_family="L2"

|||||
Db 2271 CTTCCAGGCTCAGGTGATCTCCACCTCAGCCTCCCGAGTAGCTGGACTACAGGCAT 2212
QY 297 gcatcacccgctagctattgttttacattttttgtagatgggtctcaccatgttg 356
|||||
Db 2211 GCACACACCGTCCAGCTAATTTTGTACTATTTTGTAGACGCGGTTTACCATGTGTG 2152
QY 357 ccaggtgttgctcaaacctctgggctcaagcaatccgctcagtcacccctcccaaatg 416
|||||
Db 2151 CACAGGCTGTCTCAGACTCCTGGGCTCAAGTATCCGCTCTCGGCTCCCAAGATG 2092
QY 417 ctgggattcagcgctgagccacgggcccagggctgagtaact 459
|||||
Db 2091 CTGGGATTACAAGTGGGACCGACCGTGCCTGGGCGCCGAGATCTC 2049

RESULT 11
AL133390 71872 bp DNA linear PRI 04-JAN-2001
LOCUS Human DNA sequence from clone RPI-317G22 on chromosome 1 Contains
DEFINITION part of the KMO (kynurenine 3-monooxygenase (kynurenine
3-hydroxylase)) gene, the CHML(choroideremia-like (Rab escort
protein 2)) gene, the gene for encephalopsin, ESTs, STSS, GSSs and
a CpG Island, complete sequence.
AL133390
ACCESSION AL133390.7 GI:7406530
VERSION
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 71872)
Parker, A.
Direct Submission
Submitted (04-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Apr 3, 2000 this sequence version replaced gi:7320611.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence is
the entire insert of clone RPI-317G22 This sequence has been
finished according to sequence map criteria as follows. An attempt
is made to resolve all sequencing problems, such as compressions
and repeats, but not necessarily within known annotated repeat
sequence elements. Where the sequence is ambiguous, there is an
annotation using the 'unsure' feature key. This sequence was
generated from part of bacterial clone contigs of human chromosome
1, constructed by the Sanger Centre Chromosome 1 Mapping Group.
Further information can be found at
http://www.sanger.ac.uk/HGP/Chrl
RPI-317G22 is from the library RPI-1 constructed by the group of
Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2.
Location/Qualifiers
1. 71872
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RPI-317G22"
/clone_lib="RPI-1"
repeat_region 21..2548
/note="L1PBb repeat: matches 30..2559 of consensus"
repeat_region 2509..2919

/note="L1PBa repeat: matches 601..1014 of consensus"
2920..3281
/note="THE1B repeat: matches 1..364 of consensus"
3282..4433
/note="L1PBa repeat: matches -1481..601 of consensus"
4539..7775
/note="L1M1 repeat: matches -68..3127 of consensus"
7784..7869
/note="MER2 repeat: matches 130..209 of consensus"
8165..8424
/note="L1PA5 repeat: matches 5884..6143 of consensus"
8425..8631
/note="L1P2 repeat: matches 5392..5600 of consensus"
9118..9406
/note="AluSx repeat: matches 1..286 of consensus"
complement(9535..10007)
/note="match: GSS: Em:AQ882365"
9650..9956
/note="AluSx repeat: matches 1..311 of consensus"
10366..10674
/note="AluJb repeat: matches 1..307 of consensus"
10800..10849
/note="25 copies 2 mer tt 76% conserved"
join(<10853..10910,12922..13004,14186..14287,14378..14437,
16127..19815)
/gene="KMO"
/note="match: cDNAs: Em:AF056031 Em:AF056032 Em:Y13153
match: ESTs: Em:A1074145 Em:R64497 Em:A1341967 Em:H47435
Em:R65593"
/product="dJ317G22.1 (kynurenine 3-monooxygenase
(kynurenine 3-hydroxylase))"
/evidence="not_experimental"
join(10853..10910,12922..13004,14186..14287,14378..14437,
16127..19815)
/gene="KMO"
join(<10853..10910,12922..13004,14186..14287,14378..14437,
16127..16327)
/gene="KMO"
/note="match: proteins: Tr:O15229 Tr:O88867 Tr:Q21795
Tr:O76150"
/codon_start=1
/evidence="not_experimental"
/product="dJ317G22.1 (kynurenine 3-monooxygenase
(kynurenine 3-hydroxylase))"
/protein_id="CAC19783.1"
/db_xref="GI:12043384"
/translation="GFEDCLVFDELKFDNDLCLPVSFLRIPDDHAIISDLSMYN
YTEMRAHVNSSWFIQKNMERFLHAIMPESTFIPLTMTVFSRIRIHEAVQRHHQKKV
INKGLFFLGSIATISSTYLLIHMSPRFLRLRPWNWIAHFRNTTCPPAKAVDSLEQ
ISNLISR"
repeat_region 11252..11551
/note="MER1B repeat: matches 36..329 of consensus"
11641..11916
/note="AluSx repeat: matches 2..292 of consensus"
complement(11977..12463)
/note="match: GSS: Em:AQ358898"
complement(12102..12365)
/note="match: STS: Em:G38435"
12194..12418
/note="MIR repeat: matches 33..258 of consensus"
12484..12929
/gene="KMO"
/note="match: GSS: Em:AQ814890"
12485..12534
/note="25 copies 2 mer tt 84% conserved"
complement(16125..16461)
/note="match: STS: Em:G28193"
complement(16150..16461)
/note="match: STS: Em:G25963"
complement(16804..17103)
/note="match: STS: Em:G07828"
16807..16985
/gene="KMO"

COMMENT

* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

misc_feature	1. .2950	/note="assembly_fragment"
misc_feature	3051. .6314	/note="assembly_fragment"
misc_feature	6415. .11024	/note="assembly_fragment"
misc_feature	11925. .20145	/note="assembly_fragment"
misc_feature	20246. .30884	/note="assembly_fragment"
misc_feature	30985. .44984	/note="assembly_fragment"
	clone_end.T7	vector_side:right"
misc_feature	45085. .77877	/note="assembly_fragment"
misc_feature	77978. .112578	

RESULT 13

AL365215	AL365215	184703 bp	DNA	linear	PRI 23-JUN-2001			
LOCUS	Human DNA sequence from clone RP11-416D8 on chromosome 10, complete sequence.							
ACCESSION	AL365215							
VERSION	AL365215.23	GI:14529846						
KEYWORDS	HTG.							
SOURCE	human.							
ORGANISM	Homo sapiens							
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.							
REFERENCE	1 (bases 1 to 184703)							
AUTHORS	Ashwell, R.							
TITLE	Direct Submission							
JOURNAL	Submitted (23-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk							
	requests: clonequest@sanger.ac.uk							
	clone: this sequence version replaced at: 14456224.							

COMMENT

On Jun 22, 2001 this sequence version replaced [gi:14456224](#). During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: En:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10

Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr10>
RP11-416D8 is from the library RP11-11.2 constructed by the group
of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
VECTOR: pBAC3.6

This sequence is the entire insert of clone RP11-416D8 The true
left end of clone RP11-320B8 is at 101956 in this sequence.

FEATURES

Source

1. 184703
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-416D8"
/clone.lib="RP11-11.2"

repeat_region 781..1138
/note="MLT1A2 repeat: matches 3..374 of consensus"
repeat_region 1378..1520
/note="MIR repeat: matches 13..169 of consensus"
repeat_region 1700..2027
/note="AluX repeat: matches 1..311 of consensus"
repeat_region 3259..3643
/note="MLT1B repeat: matches 1..390 of consensus"
repeat_region 4028..4397
/note="THE1C repeat: matches 1..371 of consensus"
repeat_region 4934..5497
/note="L1ME repeat: matches 866..1405 of consensus"
repeat_region 5955..6285
/note="L1MEC repeat: matches 1813..2148 of consensus"
repeat_region 6335..6655
/note="L1M4 repeat: matches 2690..3056 of consensus"
repeat_region 6656..6959
/note="AluY repeat: matches 1..299 of consensus"
repeat_region 6960..7059
/note="L1M4 repeat: matches 3056..3146 of consensus"
repeat_region 7199..7937
/note="L1PAB repeat: matches 5420..6156 of consensus"
repeat_region 7950..8103
/note="AluY repeat: matches 158..311 of consensus"
repeat_region 8106..8401
/note="AluSp repeat: matches 1..304 of consensus"
repeat_region 8420..8744
/note="L1M4 repeat: matches 3412..3788 of consensus"
repeat_region 8760..9851
/note="L1M4 repeat: matches 3904..5041 of consensus"
repeat_region 9847..10012
/note="L1MB7 repeat: matches 5979..6149 of consensus"
repeat_region 10161..10335
/note="L2 repeat: matches 2019..2198 of consensus"
repeat_region 11843..11895
/note="MIR repeat: matches 85..145 of consensus"
repeat_region 12341..12649
/note="AluJo repeat: matches 1..309 of consensus"
repeat_region 13711..13861
/note="MIR repeat: matches 99..257 of consensus"
repeat_region 13868..14232
/note="THE1B repeat: matches 1..364 of consensus"
repeat_region 14240..14328
/note="MIR repeat: matches 1..92 of consensus"
repeat_region 15157..15178
/note="11 copies 2 mer tt 100% conserved"
repeat_region 15659..16253
/note="L2 repeat: matches 1536..2155 of consensus"
repeat_region 16254..16535
/note="AluJo repeat: matches 13..293 of consensus"
repeat_region 16536..16584
/note="L2 repeat: matches 2155..2201 of consensus"
repeat_region 16585..17051
/note="MLT1D repeat: matches 24..505 of consensus"
repeat_region 17056..17346
/note="L2 repeat: matches 2196..2489 of consensus"
repeat_region 17471..17519
/note="L2 repeat: matches 2660..2708 of consensus"

repeat_region 18417..18641
/note="MIR repeat: matches 14..262 of consensus"
repeat_region 19232..19341
/note="MER5B repeat: matches 1..109 of consensus"
repeat_region 19355..19657
/note="AluJb repeat: matches 1..303 of consensus"
repeat_region 19935..20010
/note="MER3 repeat: matches 5..82 of consensus"
repeat_region 20104..20153
/note="L2 repeat: matches 2627..2676 of consensus"
repeat_region 22267..22495
/note="AluJo repeat: matches 57..293 of consensus"
repeat_region 23607..23909
/note="AluJb repeat: matches 5..306 of consensus"
repeat_region 23950..24341
/note="L1M8 repeat: matches 5899..6280 of consensus"
repeat_region 24645..24839
/note="L1MB2 repeat: matches 5962..6164 of consensus"
repeat_region 24853..25290
/note="Tigger2a repeat: matches 1..434 of consensus"
repeat_region 25291..25478
/note="94 copies 2 mer tg 61% conserved"
repeat_region 25485..25552
/note="34 copies 2 mer gt 66% conserved"
repeat_region 25491..25610
/note="3 copies 40 mer 79% conserved"
repeat_region 26969..27179
/note="MIR repeat: matches 20..232 of consensus"
repeat_region 27709..28003
/note="AluJb repeat: matches 1..288 of consensus"
repeat_region 28719..29028
/note="AluSq repeat: matches 1..311 of consensus"
repeat_region 29426..29730
/note="AluX repeat: matches 4..307 of consensus"
repeat_region 29738..29855
/note="L2 repeat: matches 2546..2654 of consensus"
repeat_region 30066..30366
/note="AluY repeat: matches 1..301 of consensus"
repeat_region 30370..30691
/note="L2 repeat: matches 2381..2709 of consensus"
repeat_region 31249..31513
/note="AluSc repeat: matches 25..306 of consensus"
repeat_region 31709..32120
/note="MSTA repeat: matches 1..424 of consensus"
repeat_region 33825..33856
/note="16 copies 2 mer ac 87% conserved"
repeat_region 34925..35218
/note="AluSq repeat: matches 1..295 of consensus"
repeat_region 35223..35538
/note="AluSq repeat: matches 1..309 of consensus"
repeat_region 37794..37990
/note="MER3 repeat: matches 2..190 of consensus"
repeat_region 38000..38172
/note="MER5A repeat: matches 11..189 of consensus"
repeat_region 39449..39470
/note="11 copies 2 mer aa 100% conserved"
repeat_region 41110..41165
/note="28 copies 2 mer aa 75% conserved"
repeat_region 41416..41715
/note="AluJb repeat: matches 1..308 of consensus"
repeat_region 42825..43079
/note="AluX repeat: matches 44..301 of consensus"
repeat_region 43309..43558
/note="AluJ/FLAM repeat: matches 1..310 of consensus"
repeat_region 45407..45480
/note="Alu repeat: matches 241..301 of consensus"
repeat_region 46510..46702
/note="AluJo repeat: matches 86..302 of consensus"
repeat_region 48748..49033
/note="AluJo repeat: matches 1..301 of consensus"
repeat_region 50196..50327
/note="FLAM_C repeat: matches 1..132 of consensus"
repeat_region 51396..51701

repeat_region /note="AluJb repeat: matches 1. .306 of consensus" 51850. .52144

repeat_region /note="AluSc repeat: matches 1. .295 of consensus" 52148. .52179

repeat_region /note="16 copies 2 mer tt 87% conserved" 52289. .52417

repeat_region /note="L1LMCA repeat: matches 7628. .7769 of consensus" 52579. .52808

repeat_region /note="AluJb repeat: matches 85. .296 of consensus" 52845. .53026

repeat_region /note="MSTA repeat: matches 221. .425 of consensus" 53027. .53391

repeat_region /note="THE1B repeat: matches 1. .363 of consensus" 53392. .53493

repeat_region /note="MSTA repeat: matches 126. .221 of consensus" 53494. .53900

repeat_region /note="MUT2B repeat: matches 1. .409 of consensus" 53494. .53900

Query Match 17.1%; Score 195; DB 9; Length 184703;
Best Local Similarity 80.6%; Pred. No. 8.6e-44;
Matches 228; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Qy 177 tttgtctgtccaccaggctgaagtacagtgttaagtcacagctcagcgtcaagctttga 236
Db 19384 TCTTGTCTGTCAACCCAGCAGGAGTCAGTGGCAGCATCACTGCTCACTGCGAGCTTGA 19443

Qy 237 cctccaggctcaagtatctctcgtcagctttccaagtaactgggacacagcgtcat 296
Db 19444 CCTCCAGGCTCAGGTATCCTCCACTCAGCCTCCCGAGTAGCTGGGACTACAGGCAT 19503

Qy 297 gcatcaccacagctaggctattgttttacattttgttagatggggtctcacacattgtg 356
Db 19504 GCACCACCACTCCAGCTATTTTGTACTTTTGTAGACGGGCTTCCACCATGTTG 19563

Qy 357 cccagggtgtctcaactcctgggtcgaagcaatccgcgtcaagctcaccctcccaaatg 416
Db 19564 CACAGGCTGCTCAGACTCGGGCTCAAGTATCGCCCTGCTTGGCCCTCCCAAGTG 19623

Qy 417 ctgggattacaggtgaagcaccgcccaggtcgtgaatc 459
Db 19624 CTGGGATTACAACTGGGAGCCACCGCTGCCCGCCCGAGAYCTC 19666

RESULT 14
AC090104/c AC090104 164126 bp DNA linear HTG 09-MAY-2001
DEFINITION Homo sapiens chromosome 8 clone RP11-451E11 map 8, WORKING DRAFT
SEQUENCE, 7 unordered pieces.
AC090104 AC090104 GI:13959193
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 164126)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 8, clone RP11-451E11
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 164126)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Choepel,Y., Collangelo,M., Collins,S., Collumore,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galaagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Karats,A., LaRocque,K., Lamazares,R., Landers,I., Lehoczy,J., Levine,R., Liu,G., MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,

Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission
Submitted (14-FEB-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 7, 2001 this sequence version replaced gi:13940698.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L12770
Center clone name: 451_E11
----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 161850 bases at least Q40
Consensus quality: 162799 bases at least Q30
Consensus quality: 163176 bases at least Q20
Insert size: 163000; agarose-fp
Insert size: 163526; sum-of-contigs
Quality coverage: 11.8 in Q20 bases; agarose-fp
Quality coverage: 11.7 in Q20 bases; sum-of-contigs

***** NOTE: This is a 'working draft' sequence. It currently consists of 7 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 21292: contig of 21292 bp in length
21293 21392: gap of 100 bp
21393 22889: contig of 1497 bp in length
22890 22989: gap of 100 bp
22990 26384: contig of 3395 bp in length
26385 26484: gap of 100 bp
26485 95293: contig of 68809 bp in length
95294 95393: gap of 100 bp
95394 119296: contig of 23903 bp in length
119297 119396: gap of 100 bp
119397 147115: contig of 27719 bp in length
147116 147215: gap of 100 bp
147216 164126: contig of 16911 bp in length.

FEATURES
source
1. .164126
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-451E11"
/clone_lib="RP11-451E11 Human Male BAC"
misc_feature
1. .21292
/note="assembly_fragment
vector_side:left"
21393..22889
/note="assembly_fragment"
22990..26384
/note="assembly_fragment"
26485..95293
misc_feature
misc_feature
misc_feature
misc_feature

Db 156315 CAGCCT 156310

Search completed: September 24, 2002, 15:48:01
Job time: 19976 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2002, 11:33:25 ; Search time 403.26 Seconds
(without alignments)
4857.907 Million cell updates/sec

Title: US-09-617-174B-1
Perfect score: 1141
Sequence: 1 agatagcacagcagagag.....aggataactgtgactccagg 1141

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	1137.8	99.7	1141	Human maspin promo
2	745.4	65.3	12426	Tumour suppressor
3	723.4	63.4	12426	Tumour suppressor
4	195	17.1	568	Human cDNA clone (
5	195	17.1	2343	Human cDNA sequenc
6	185.4	16.2	916	Human reproductive
7	185.4	16.2	32184	Human nervous syst
8	185.4	16.2	32184	Human immune syst
9	185.4	16.2	36272	Human immune/haema

c 10	185.4	16.2	38348	22	AAK84953	Human immune/haema
c 11	184.6	16.2	240825	22	AAF24497	Human PG-3 gene.
c 12	184	16.1	1162	21	AAF79682	Human secreted pro
c 13	183.4	16.1	735	22	AAK05557	Human reproductive
c 14	183.4	16.1	735	22	AAK09703	Human digestive sy
c 15	182.6	16.0	1956	22	ABA21277	Human nervous syst
c 16	182.2	16.0	17026	22	AAK70372	Human immune/haema
c 17	181.2	15.9	1519	20	AA210674	cDNA encoding a hu
c 18	180.8	15.8	1421	22	AAK73297	Human immune/haema
c 19	180.4	15.8	21477	22	AAK66626	Human immune/haema
c 20	180.4	15.8	21480	22	AAK66625	Human immune/haema
c 21	179.8	15.8	1423	22	AAF27785	Human MANGO 347 co
c 22	179.8	15.8	1440	21	AAF30029	Clone HTHCW70 codi
c 23	179.8	15.8	1442	22	AAH33267	Human colon cancer
c 24	179.2	15.7	5269	22	AAK80329	Human immune/haema
c 25	179	15.7	7233	22	AAK36623	Human cardiovascular
c 26	178.8	15.7	8053	22	AAI98931	Human excretory re
c 27	178.8	15.7	8053	22	AAI63281	Human kidney relat
c 28	178.8	15.7	10241	22	AAK83768	Human immune/haema
c 29	178.8	15.7	10241	22	AAK33394	DNA encoding human
c 30	178.6	15.7	1125	22	ABA06543	Human cDNA SEQ ID
c 31	178.6	15.7	1125	22	AAK28831	Human immunoglobul
c 32	178.4	15.6	16181	22	AAK04268	Human reproductive
c 33	178	15.6	5962	23	AAK82348	DNA encoding novel
c 34	178	15.6	5962	23	AAK85934	DNA encoding novel
c 35	177.8	15.6	700	20	AAK30339	DNA encoding a hum
c 36	177.6	15.6	4071	22	AAK36540	Human cardiovascular
c 37	177.6	15.6	4071	22	AAK85200	Human immune/haema
c 38	177.2	15.5	21458	22	AAK06122	Human reproductive
c 39	177.2	15.5	21458	22	AAK81266	Human immune/haema
c 40	177	15.5	596	22	AAK75914	Human immune/haema
c 41	177	15.5	596	22	AAK75914	Human immune/haema
c 42	176.8	15.5	4431	22	ABA18399	Human nervous syst
c 43	176.8	15.5	4431	22	AAK40513	DNA encoding human
c 44	176.8	15.5	4431	22	AAK04139	Human reproductive
c 45	176.8	15.5	4433	22	ABA18400	Human nervous syst

ALIGNMENTS

RESULT 1
AAT89000
ID AAT89000 standard; cDNA: 1141 BP.

XX AC AAT89000;

XX DT 22-APR-1998 (first entry)

XX DE Human maspin promoter and partial coding cDNA.

XX KW Maspin; serpin; mammary epithelial cell; human; promoter; malignant;
tumour cell; treatment; prostate cancer; breast cancer; metastasis; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers
FT promoter 1..1141 a

FT FT /tag= a
/note= "maspin promoter region"

FT FT misc_signal 443..449
/tag= b

FT FT misc_signal 451..457
/note= "Putative regulatory element AP2"

FT FT misc_signal 485..490
/note= "Putative regulatory element AP1"

FT FT misc_signal 684..691
/tag= d

FT FT misc_signal 684..691
/note= "Putative regulatory element Ets"

FT FT misc_signal 846..851
/tag= e

FT FT misc_signal 846..851
/note= "Putative regulatory element HRE"

XX PD 20-SEP-2001.

XX PF 15-MAR-2001; 2001WO-EP02955.

XX PR 15-MAR-2000; 2000DE-1013847.

XX PR 06-APR-2000; 2000DE-1019058.

XX PR 07-APR-2000; 2000DE-1019173.

XX PR 30-JUN-2000; 2000DE-1032529.

XX PR 01-SEP-2000; 2000DE-1043826.

XX PA (EPIC-) EPIGENOMICS AG.

XX PI Olek A, Piepenbrock C, Berlin K;

XX DR WPI; 2001-602752/68.

XX PT Fragments of chemically modified genes associated with tumour suppressor

XX PT genes and oncogenes, useful in designing primers and probes for

XX PT analysing diseases associated with cytosine methylation state e.g.

XX PT cancer -

XX PS Claim 1; SEQ ID No 204; 27pp; English.

XX CC The invention relates to a nucleic acid comprising a sequence of 18

XX CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with

XX CC bisulphite, of genes associated with tumour suppression and

XX CC oncogenes having a sequence taken from 536 (actually 533 since

XX CC numbers 408, 458 and 500 are missing from the sequence listing) sequences

XX CC (SS) and sequences complementary to (SS). The nucleic acid may be a

XX CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may

XX CC form part of a set of probes for detecting the cytosine methylation state

XX CC and/or single nucleotide polymorphisms and also to be used in an

XX CC array for analysing diseases associated with CpG dinucleotides e.g.

XX CC cancers and tumours. The probes can also be used in a method for

XX CC ascertaining genetic and/or epigenetic parameters for the diagnosis

XX CC and/or therapy of existing diseases or the predisposition to specific

XX CC diseases, by analysing cytosine methylations. The parameters may be

XX CC compared to another set of genetic and/or epigenetic parameters, the

XX CC differences serving as basis for diagnosis and/or prognosis events which

XX CC are disadvantageous to patients. The present sequence is one of the

XX CC 533 genomic sequences derived from tumour suppressor genes and

XX CC oncogenes. Sequences with even numbered Seq ID numbers are the

XX CC complementary sequence of the corresponding odd numbered sequence (e.g.

XX CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence

XX CC is missing).

XX CC Note: The sequence data for this patent did not form part

XX CC of the printed specification, but was obtained in electronic

XX CC format directly from WIPO at

XX CC ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 12426 BP; 3791 A; 137 C; 2517 G; 5981 T; 0 other;

Query Match 65.3%; Score 745.4; DB 22; Length 12426;

Best Local Similarity 78.4%; Pred. No. 1.9e-193;

Matches 893; Conservative 0; Mismatches 246; Indels 0; Gaps 0;

QY 1 agataagcacagcagagaacacagctccgtttcaggtctcttctgagcgtgattcgcg 60

DB 8506 AATAAACACACAAAAAACAACACCACTCCCGTTTCAAAATCTTTCTAAAACTAAATTCG 8447

QY 61 gctggagggagtagtccaccacaaatgagaagctgtggaagcagagagacagaac 120

DB 8446 ACTAAAAAATAATCCCGCCCAATAAAAAAATAAAAAAACAACAAAAACAAC 8387

QY 121 aggtccacagagagatccagagcagagctgcgtactcttcttcttcttcttctt 180

DB 8386 AAACTCCAGCAAAAAATTCAAAACAACAACTACGTACTCTCTTTCTTTTATTCTTTT 8327

QY 181 gctcttcacccagcgtgaagtacagtggttagctcacgcgtcagctgagcttgacctc 240

DB 8326 ACTCTATCACCCCAAACTAAATAATTAACCTCAGCACTCACTACAACTTTAACTCCTC 8267

QY 241 ccaggctcaagtgatcctctcgtctcagcttttccaaagtaactgaggaccacagcagcatgc 300

DB 8266 CCAAACTCAAAATAATCTCTCGTCTCAACTTTTCCAAATAACTAAACACCAACATACAT 8207

QY 301 caccacgctagcgtatctgttttacatttttttagagagatgaggggtctcaccatgtctccca 360

DB 8206 CACCACACTAAACATATATTTTACATTTTATATAAAATAAAATCTCACCATATTAACCA 8147

QY 361 ggttggtctcaaaactccttggtcctcaagcaatccgctcagctcaacccctccccaaatcgtcg 420

DB 8146 AATTAAATCTCAAACTCTTAACCTCAACCAATCCGCTCAGCTCAACCTCCCAATAACTAA 8087

QY 421 gattacagcgtgagccaccggtggtgagtaactcctaataatcacaagagattttaaata 480

DB 8086 AATTACAAACGTAACACCCGCGCCAAACCTAAATAATCTTAATCACAAAATTTTAAAAA 8027

QY 481 gaaactcctgcgcacccaccattaaacaatatctcctaccacatttggtagtaataatttgc 540

DB 8026 AAAAATCTTCCACGCCACCCATTAACCAATATCTCTTACCAATTTAATAATAATATTTTA 7967

QY 541 ctaaatagttacctaatttttaggtagcactgtgtttatcatatatcatatcctctctttt 600

DB 7966 CTAAATAATACCTAAATTTTAAATAAACACTATATTTATATATATATATCCATTTCTCTTT 7907

QY 601 ttgattgtcttctgttttaagtggcagctacctctcttggcatctgacagaaatgagctgc 660

DB 7906 TTAATTTATCTTCTATTTAATAAACAACTACCTCTCTTAACATCTAACAAAAATAAACTAC 7847

QY 661 tgcagtttacacaaaagaatggagatcagagtaactttttgtgccaccacagtgctctgag 720

DB 7846 TACAATTTTACACAAAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATA 7787

QY 721 aaattgtagtttactatcatcacataacttcttttttttttttttttttttttttttttt 780

DB 7786 AAATTTATATATATTTACTATCATCATACATTTACTTTTATTTTCATCGAATATTTTCACTTCC 7727

QY 781 ggtcctcgtcgtggcagagagattgctgacgcatgctgacgcatgctgacgcatgacgcat 840

DB 7726 GATCTCTACGTAAACCCGCAAAAAATACCGTCAGCATATCTATACGTATACATATAACTAC 7667

QY 841 agccctctcctgcgcgaacatgctggagagccttttggagccttggagccttggagccttgc 900

DB 7666 AACCCCTCTCTACCGCAACATATATAAAACCTTTTAAAAAATAATATAAAATAAAATAAA 7607

QY 901 tcagcctgaatcatctcttctcaattgtgacaagctgccaagagccttgaagtaggagag 960

DB 7606 TCAACCTAAATCATTTCTTTCAATTTATAACAAACCTACCAAAAAAATAATATAAAATAAA 7547

QY 961 agtgcctcgcagagcgtggcgtggcgtggcgtggcgtggcgtggcgtggcgtggcgtgg 1020

DB 7546 AATACCGCGCAACGAAACGAAACGAAACGAAACGAAACGAAACGAAACGAAACGAAACG 7487

QY 1021 gctgcccaggtgagccaccgctgctcttctgcccagacagcgtgctcctccacatccaggtct 1080

DB 7486 ACTACCAAAATAAACCAACCGCTACTTCTTACCCAAACGATCGCTCCACATCAAAATCT 7427

QY 1081 ttgtgctcctcgtctgctgctcttcttctccagcatttccagagataaactgtgaactcca 1139

DB 7426 TTATACTCTCTCGTTCCTTATCTTCTTTTCCACGCAATTTTCCAAAAATAAACTATAA 7368

RESULT 3

AAS46481

ID AAS46481 standard; DNA: 12426 BP.

XX

AC AAS46481;

XX

DT 18-DEC-2001 (first entry)

XX

DE Tumour suppressor gene derived chemically modified sequence #203.

XX Human; tumour suppressor gene; oncogene; antitumour; cytostatic;

KW

	-	9
	-	8
	-	7
	-	6
	-	5
	-	4
	-	3
	-	2
	-	1
	-	0
	-	9
	-	8
	-	7
	-	6
	-	5
	-	4
	-	3
	-	2
	-	1
	-	0
	-	9
	-	8
	-	7
	-	6
	-	5
	-	4
	-	3
	-	2
	-	1
	-	0
	-	9
	-	8
	-	7
	-	6
	-	5
	-	4
	-	3
	-	2
	-	1
	-	0
	-	9
	-	8
	-	7
	-	6
	-	5
	-	4
	-	3
	-	2
	-	1
	-	0
	-	9
	-	8
	-	7
	-	6
	-	5
	-	4
	-	3
	-	2
	-	1
	-	0
	-	9
	-	8
	-	7
	-	6
	-	5
	-	4
	-	3
	-	2
	-	1
	-	0
	-	9
	-	8
	-	7
	-	6
	-	5
	-	4
	-	3
	-	2
	-	1
	-	0
	-	9
	-	8
	-	7
	-	6
	-	5
	-	4
	-	3
	-	2
	-	1
	-	0
	-	9
	-	8
	-	7
	-	6
	-	5
	-	4
	-	3
	-	2
	-	1
	-	0
	-	9
	-	8
	-	7
	-	6
	-	5
	-	4
	-	3
	-	2
	-	1
	-	0
	-	9
	-	8
	-	7
	-	6
	-	5
	-	4
	-	3
	-	2
	-	1
	-	0
	-	9
	-	8
	-	7
	-	6
	-	5
	-	4
	-	3
	-	2
	-	1
	-	0
	-	9
	-	8
	-	7
	-	6
	-	5
	-	4
	-	3
	-	2
	-	1
	-	0
	-	9
	-	8
	-	7
	-	6
	-	5
	-	4
	-	3
	-	2
	-	1
	-	0
	-	9
	-	8
	-	7
	-	6
	-	5
	-	4
	-	3
	-	2
	-	1
	-	0
	-	9
	-	8
	-	7
	-	6
	-	5
	-	4
	-	3
	-	2
	-	1
	-	0
	-	9
	-	8
	-	7
	-	6
	-	5
	-	4
	-	3
	-	2
	-	1
	-	0
	-	9
	-	8
	-	7
	-	6
	-	5
	-	4
	-	3
	-	2
	-	1
	-	0
	-	9
	-	8
	-	7
	-	6
	-	5
	-	4
	-	3
	-	2
	-	1
	-	0
	-	9
	-	8
	-	7
	-	6
	-	

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234597.
PR 25-SEP-2000; 2000US-0234598.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.

PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249266.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
PA Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-541565/60.
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases
XX Disclosure; SEQ ID NO 12920; 1701pp + Sequence Listing; English.
XX The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 32184 BP; 7866 A; 6152 C; 6798 G; 11368 T; 0 other;

Query Match 16.2%; Score 185.4; DB 22; Length 32184;
Best Local Similarity 74.3%; Pred. No. 5.8e-40;

XX	17-JAN-2001;	2001WO-US01354.
PF		
XX		
PR	31-JAN-2000;	2000US-0179065.
PR	04-FEB-2000;	2000US-0180628.
PR	04-FEB-2000;	2000US-0184626.
PR	02-MAR-2000;	2000US-0186350.
PR	16-MAR-2000;	2000US-0189874.
PR	17-MAR-2000;	2000US-0190076.
PR	18-APR-2000;	2000US-0198123.
PR	19-MAY-2000;	2000US-0205515.
PR	07-JUN-2000;	2000US-0209467.
PR	28-JUN-2000;	2000US-0214886.
PR	30-JUN-2000;	2000US-0215135.
PR	07-JUL-2000;	2000US-0216647.
PR	11-JUL-2000;	2000US-0216880.
PR	11-JUL-2000;	2000US-0217487.
PR	11-JUL-2000;	2000US-0217496.
PR	14-JUL-2000;	2000US-0218290.
PR	26-JUL-2000;	2000US-02192963.


```

FT primer_bind /*tag= i 39954..39972
FT 4559..4577
FT /*tag= ao
FT 39961..39985
FT misc_binding /*tag= ap
FT /*note= "binds probe"
FT 4589..4613
FT /*tag= l
FT /*note= "binds probe"
FT 4601..4620
FT /*tag= m
FT primer_bind complement(4602..4620)
FT 4627..4718
FT /*tag= n
FT exon
FT /*tag= o
FT primer_bind 4891..4908
FT /*tag= p
FT 10007..10025
FT primer_bind /*tag= q
FT 10115..10233
FT exon
FT /*tag= r
FT 10209..10227
FT primer_bind /label= "B"
FT 10216..10240
FT misc_binding /*tag= s
FT /*tag= t
FT /*note= "binds probe"
FT 10228..10240
FT allele replace(10228,T)
FT /*tag=
FT primer_bind complement(10229..10247)
FT 10267..10285
FT primer_bind /*tag= v
FT /*tag= w
FT misc_binding 10274..10298
FT /*tag= x
FT /*note= "binds probe"
FT 10286..10298
FT allele replace(10286,T)
FT /*tag= y
FT primer_bind complement(10287..10305)
FT /*tag= z
FT allele replace(10370,)
FT /*tag= aa
FT 10411..10430
FT primer_bind /*tag= ab
FT 26810..26897
FT exon /*tag= ac
FT /*label= "D"
FT 31357..31471
FT exon /*tag= ad
FT /*label= "E"
FT 34261..34404
FT exon /*tag= ae
FT /*label= "F"
FT 37377..37466
FT exon /*tag= af
FT /*label= "S"
FT 39704..40858
FT exon /*tag= ag
FT /*label= "T"
FT 39556..39574
FT primer_bind /*tag= ah
FT 39877..39896
FT primer_bind /*tag= ai
FT 39925..39943
FT primer_bind /*tag= aj
FT 39932..39956
FT misc_binding /*tag= ak
FT /*note= "binds probe"
FT 39944..39963
FT allele replace(39944,T)
FT /*tag= al
FT primer_bind complement(39945..39963)
FT 39953..39970
FT primer_bind

```

```

FT primer_bind /*tag= an
FT 39954..39972
FT /*tag= ao
FT 39961..39985
FT misc_binding /*tag= ap
FT /*note= "binds probe"
FT 41373..41384
FT allele replace(39973,C)
FT /*tag= aq
FT primer_bind complement(39974..39992)
FT /*tag= ar
FT 40242..40259
FT primer_bind /*tag= as
FT 41137..41154
FT /*tag= at
FT 41366..41384
FT primer_bind /*tag= au
FT 41373..71397
FT misc_binding /*tag= av
FT /*note= "binds probe"
FT 41385..41403
FT allele replace(41385,C)
FT /*tag= aw
FT 41385..41403
FT primer_bind /*tag= ax
FT 41385..41403
FT /*tag= ay
FT 41392..41416
FT misc_binding /*tag= az
FT /*note= "binds probe"
FT 41404..41423
FT allele replace(41404,C)
FT /*tag= ba
FT 41564..41581
FT primer_bind /*tag= bb
FT 42122..42141
FT /*tag= bc
FT 42122..42141
FT primer_bind /*tag= bd
FT 42213..42231
FT /*tag= be
FT 42220..42244
FT misc_binding /*tag= bf
FT /*note= "binds probe"
FT 42232..42251
FT allele replace(42232,C)
FT /*tag= bg
FT 42233..42251
FT primer_bind complement(42233..42251)
FT /*tag= bh
FT 42526..42543
FT primer_bind /*tag= bi
FT 50436..50545
FT exon /*tag= bj
FT /*label= "G"
FT 67289..67309
FT primer_bind /*tag= bk
FT 67456..67474
FT primer_bind /*tag= bl
FT 67463..67487
FT misc_binding /*tag= bm
FT /*note= "binds probe"
FT 67475..67494
FT allele replace(67475,G)
FT /*tag= bn
FT 67476..67494
FT primer_bind complement(67476..67494)
FT /*tag= bo
FT 67724..67741
FT primer_bind /*tag= bp
FT 69182..69200
FT /*tag= bq
FT 69502..69520
FT primer_bind /*tag= br
FT 69509..69533
FT misc_binding /*tag= bs
FT /*note= "binds probe"
FT 69521..69540
FT allele replace(69521,G)
FT /*tag= bt
FT 69522..69540
FT primer_bind complement(69522..69540)

```


XX AAL05557;
XX 21-NOV-2001 (first entry)
XX Human reproductive system related antigen DNA SEQ ID NO: 8245.
XX Human; reproductive system related antigen; reproductive system disorder;
XX cancer; gene therapy; ds.
XX Homo sapiens.
XX WO20015320-A2.
XX 02-AUG-2001.
XX 17-JAN-2001; 2001WO-US01339.
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214986.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.
XX 22-AUG-2000; 2000US-0226868.
XX 22-AUG-2000; 2000US-0227182.
XX 23-AUG-2000; 2000US-0227009.
XX 30-AUG-2000; 2000US-0228924.
XX 01-SEP-2000; 2000US-0229287.
XX 01-SEP-2000; 2000US-0229343.
XX 01-SEP-2000; 2000US-0229344.
XX 01-SEP-2000; 2000US-0229345.
XX 05-SEP-2000; 2000US-0229509.
XX 05-SEP-2000; 2000US-0229513.
XX 06-SEP-2000; 2000US-0230437.
XX 06-SEP-2000; 2000US-0230438.
XX 08-SEP-2000; 2000US-0231242.
XX 08-SEP-2000; 2000US-0231243.
XX 08-SEP-2000; 2000US-0231244.
XX 08-SEP-2000; 2000US-0231413.
XX 08-SEP-2000; 2000US-0231414.
XX 08-SEP-2000; 2000US-0232080.
XX 08-SEP-2000; 2000US-0232081.
XX 12-SEP-2000; 2000US-0231968.
XX 14-SEP-2000; 2000US-0232397.
XX 14-SEP-2000; 2000US-0232398.
XX 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235835.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.

PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216680.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.

PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0242221.
PR 08-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251866.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-541565/60.

Nucleic acids encoding 3224 human nervous system antigen polypeptides, useful for preventing, diagnosing and/or treating nervous system cancers and metastases -

Disclosure; SEQ ID NO 13608; 170lpp + Sequence Listing; English.

The invention relates to novel genes (ABAl1004-ABAl21534) and proteins (ABAl14678-ABAl8001) useful for preventing, treating or ameliorating

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2002, 10:07:45 ; Search time 3309.28 seconds
(without alignments)
4653.588 Million cell updates/sec

Title: US-09-617-174B-1

Perfect score: 1141

Sequence: 1 agataacacagcagagaag.....aggataactgtgactccagg 1141

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST: *

1: em_estba: *
2: em_esthum: *
3: em_estin: *
4: em_estmu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_htc: *
9: gb_est1: *
10: gb_est2: *
11: gb_htc: *
12: gb_gss: *
13: em_gss_hum: *
14: em_gss_inv: *
15: em_gss_pln: *
16: em_gss_vrt: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	197.8	17.3	496	12 B35967	B35967 HS-1031-A2-
c 2	195	17.1	568	9 AU159406	AU159406 AU159406
c 3	186.2	16.3	410	9 AU1023375	AU1023375 ow63902.x
c 4	184.2	16.1	708	12 AG143040	AG143040 Pan trogl
c 5	184.2	16.1	713	12 AQ312217	AQ312217 RPCI11-10
c 6	183.6	16.1	655	12 AG169306	AG169306 Pan trogl
c 7	183	16.0	643	9 AU114719	AU114719 HAI1318 Hu
c 8	183	16.0	2167	11 AF116633	AF116633 Homo sapi
c 9	182.4	16.0	390	10 R56162	R56162 y991b06.r1
c 10	182.4	16.0	742	12 AQ343483	AQ343483 RPCI11-12
c 11	181.8	15.9	649	12 BH609847	BH609847 HIV13C04
c 12	181.6	15.9	524	9 AW973992	AW973992 EST385990
c 13	180.6	15.8	571	9 AA131088	AA131088 zol16c04.s
c 14	179.8	15.8	331	10 BG057233	BG057233 7f75d01.x
c 15	179.8	15.8	661	9 BE177530	BE177530 RCI-HT059
c 16	179.8	15.8	698	12 AG141964	AG141964 Pan trogl
c 17	179.6	15.7	430	12 AQ393433	AQ393433 CITBI-E1-

18	179.4	15.7	424	9 AI499954	AI499954 to63c10.x
19	179.2	15.7	373	9 AL596605	AL596605 DKF2p451C
20	179.2	15.7	559	10 BG537830	BG537830 602566186
21	179.2	15.7	713	10 BG684208	BG684208 602635838
22	179.2	15.7	893	10 BI913559	BI913559 603180311
c 23	179.8	15.7	1048	9 AL576528	AL576528 AL576528
c 24	178.8	15.7	2205	11 AF318346	AF318346 Homo sapi
c 25	178	15.6	455	12 AQ392348	AQ392348 CITBI-E1-
c 26	178	15.6	558	12 AQ527359	AQ527359 RPCI-11-3
c 27	177.8	15.6	379	9 AV754799	AV754799 AV754799
c 28	177.8	15.6	684	12 AG037064	AG037064 Pan trogl
c 29	177.4	15.5	456	9 AI753334	AI753334 cr08f09.x
c 30	177.4	15.5	512	12 AQ538435	AQ538435 RPCI-11-3
c 31	177.2	15.5	416	12 AQ748789	AQ748789 HS_5574_A
c 32	177.2	15.5	515	12 AQ506282	AQ506282 RPCI-11-3
c 33	177	15.5	532	12 AQ701252	AQ701252 HS_2132_A
c 34	177	15.5	632	9 AI821267	AI821267 ab90f07.x
c 35	176.6	15.5	562	12 AQ778690	AQ778690 HS_2235_A
c 36	176.6	15.5	701	12 AG148726	AG148726 Pan trogl
c 37	176.4	15.5	512	12 AQ120174	AQ120174 HS_3028_B
c 38	176.4	15.5	950	12 AQ749313	AQ749313 HS_5576_A
c 39	176.2	15.4	471	9 AW151925	AW151925 xf70d02.x
c 40	176.2	15.4	731	12 AG175563	AG175563 Pan trogl
c 41	176.2	15.4	900	10 BE880692	BE880692 601490306
c 42	176.2	15.4	1020	10 BG615396	BG615396 602643974
c 43	176	15.4	387	9 AA736485	AA736485 nv92h02.r
c 44	176	15.4	792	9 AU136365	AU136365 AU136365
c 45	175.8	15.4	463	12 AQ033859	AQ033859 HS_2194_B

ALIGNMENTS

RESULT 1

B35967/c	B35967	HS-1031-A2-F07-MR-V0001.abi	CIT Human Genomic Sperm Library C Homo sapiens genomic clone Plate=CT 811 Col=14 Row=K, DNA sequence.	496 bp	DNA	linear	GSS 17-OCT-1997
LOCUS	B35967	HS-1031-A2-F07-MR-V0001.abi	CIT Human Genomic Sperm Library C Homo sapiens genomic clone Plate=CT 811 Col=14 Row=K, DNA sequence.	496 bp	DNA	linear	GSS 17-OCT-1997
DEFINITION	B35967	HS-1031-A2-F07-MR-V0001.abi	CIT Human Genomic Sperm Library C Homo sapiens genomic clone Plate=CT 811 Col=14 Row=K, DNA sequence.	496 bp	DNA	linear	GSS 17-OCT-1997
ACCESSION	B35967	HS-1031-A2-F07-MR-V0001.abi	CIT Human Genomic Sperm Library C Homo sapiens genomic clone Plate=CT 811 Col=14 Row=K, DNA sequence.	496 bp	DNA	linear	GSS 17-OCT-1997
VERSION	B35967.1	GI:25353336					
KEYWORDS	GSS.						
SOURCE	human.						
ORGANISM	Homo sapiens						
REFERENCE	1 (bases 1 to 496)						
AUTHORS	Mahairas,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S., Traicoff,R., Abajian,C., Blanchard,A., West,A. and Hood,L.E.						
TITLE	Construction of a Characterized Clone Resource for Genomic Sequencing: Generation and Preliminary Analysis of 20,000 Sequence Tagged Connectors						
JOURNAL	Unpublished (1997)						
COMMENT	Contact: Mahairas GG, Zackrone KD, Hood L University of Washington Seattle, WA 98195, USA Tel: (206) 616-8744 Fax: (206) 685-7301 Email: kzackrone@u.washington.edu Sequence Tagged Connector Plate: CT 811 row: K column: 14 Class: BAC ends High quality sequence stop: 496.						
FEATURES	Location/Qualifiers						
source	1..496						
	/organism="Homo sapiens"						
	/db_xref="taxon:9606"						
	/clone="Plate=CT 811 Col=14 Row=K"						
	/clone.lib="CIT Human Genomic Sperm Library C"						
	/sex="M"						
	/note="Organ: sperm; Vector: pBelBAC11; BAC Clones in E-Coli DH10B"						
BASE COUNT	155 a	85 c	135 g	120 t			
ORIGIN							


```

Query Match          16.0%; Score 183; DB 9; Length 643;
Best Local Similarity 77.4%; Pred. No. 9.2e-29;
Matches 222; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 177 ttctgctgtcacccaggctgaagtacagtggttagctcacggctcactgcagctttga 236
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 100 TCTTGCTCTGTCTCACCAGGCTGGAGTGCAGTGACAAATCTTGGCTCATAATAGCCTCGA 159

QY 237 cctccaggctcaagtgcctctctgctcagctcttccaaagtaactgggaccacaggcat 296
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 160 CTTCCAGGCTCAGGTGATCTCCACCTCAGCTCCCAACTAGCTGGGACACACAGGTGC 219

QY 297 gcatcacacgctaggtctatttttaccatttttttagatagatggggtctcaccatgttg 356
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 220 CCGCCACACAGGCTGGCTATTTTGTATTTGTGTAGATAGGGGTTTGGCATGTGTG 279

QY 357 ccaggttggtctcaaacctcctgggtcgaagcaatccgctcacgtcaaacctccccaaatg 416
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 280 GCATGCTGCTCGAAGCTCTGGGCTCAAGCAATCCACCATCTCAGCTCCCAAGTG 339

QY 417 ctggattacagctgaagccaccgggcccaggctgagtaactcctaa 463
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 340 CTGGATTACAGCGTGGGCCACACACTCGGCCAACAAATTTCTTAA 386

RESULT 8
R56162 AF116633 2167 bp mRNA linear HTC 08-MAY-2001
LOCUS Homo sapiens PRO1318 mRNA, complete cds.
DEFINITION AF116633
ACCESSION AF116633.1 GI:7959767
VERSION HTC
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 2167)
AUTHORS Zhang,C., Yu,Y., Zhang,S., Wei,H., Zhou,G., Ouyang,S., Luo,L.,
          Bi,J., Liu,M. and He,F.
TITLE Functional prediction of the coding sequences of 121 new genes
          deduced by analysis of cDNA clones from human fetal liver
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2167)
AUTHORS Zhang,C., Yu,Y., Zhang,S., Wei,H., Zhou,G., Ouyang,S., Luo,L.,
          Bi,J., Liu,M. and He,F.
TITLE Direct Submission
JOURNAL Submitted (24-DEC-1998) Department of Experimental Hematology,
          Institute of Radiation Medicine, Beijing Taiping Road 27, Beijing
          100850, P. R. China
FEATURES
    Location/Qualifiers
        1..2167
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="FLB5032"
            /tissue_type="liver"
            /dev_stage="fetal"
            /complement(85..365)
            /rpt_family="Alu"
            /complement(411..543)
            /rpt_family="MER45"
            426..689
            /note="predicted protein of HQ1318"
            /codon_start=1
            /product="PRO1318"
            /protein_id="AAF71056.1"
            /db_xref="GI:7959768"
            /translation="MPSVAQGPVPHLGSRAVAFELVMFEQRPYVILHWAPQIT
            WFLRRGVSHQSPKSPLEVLNTEAFKLSSVGETVHHHTQ"
            1137..1304
            /rpt_family="Alu"
repeat_region 643 a 420 c 408 g 696 t
BASE COUNT

```

```

Query Match          16.0%; Score 183; DB 11; Length 2167;
Best Local Similarity 77.4%; Pred. No. 9.7e-29;
Matches 222; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 177 ttctgctgtcacccaggctgaagtacagtggttagctcacggctcactgcagctttga 236
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 100 TCTTGCTCTGTCTCACCAGGCTGGAGTGCAGTGACAAATCTTGGCTCATAATAGCCTCGA 159

QY 237 cctccaggctcaagtgcctctctgctcagctcttccaaagtaactgggaccacaggcat 296
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 160 CTTCCAGGCTCAGGTGATCTCCACCTCAGCTCCCAACTAGCTGGGACACACAGGTGC 219

QY 297 gcatcacacgctaggtctatttttaccatttttttagatagatggggtctcaccatgttg 356
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 220 CCGCCACACAGGCTGGCTATTTTGTATTTGTGTAGATAGGGGTTTGGCATGTGTG 279

QY 357 ccaggttggtctcaaacctcctgggtcgaagcaatccgctcacgtcaaacctccccaaatg 416
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 280 GCATGCTGCTCGAAGCTCTGGGCTCAAGCAATCCACCATCTCAGCTCCCAAGTG 339

QY 417 ctggattacagctgaagccaccgggcccaggctgagtaactcctaa 463
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 340 CTGGATTACAGCGTGGGCCACACACTCGGCCAACAAATTTCTTAA 386

RESULT 9
R56162 R56162 390 bp mRNA linear EST 23-MAY-1995
LOCUS Yg91b06.r1 Soares infant brain IN1B Homo sapiens cDNA clone
DEFINITION IMAGE:40776 5' similar to contains Alu repetitive element;; mRNA
          sequence.
ACCESSION R56162
VERSION R56162.1 GI:826268
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 390)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
          M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
          Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevasakis,E., Waterston
          R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
          Washington University School of Medicine
          4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
          Tel: 314 286 1800
          Fax: 314 286 1810
          Email: est@watson.wustl.edu
          Insert Size: 1266
          High quality sequence stops: 349 Source: IMAGE Consortium, LLNL
          This clone is available royalty-free through LLNL; contact the
          IMAGE Consortium (info@image.llnl.gov) for further information.
          Insert Length: 1266 Std Error: 0.00
          Seq primer: M13RP1
          High quality sequence stop: 349.
FEATURES
    Location/Qualifiers
        1..390
            /organism="Homo sapiens"
            /db_xref="GDB:413317"
            /db_xref="taxon:9606"
            /clone="IMAGE:40776"
            /clone_lib="Soares infant brain IN1B"
            /sex="female"
            /dev_stage="73 days post natal"
            /lab_host="DH10B (ampicillin resistant)"
            /note="Organ: whole brain; Vector: Lafmid BA; Site:1: Not
            1; Site_2: Hind III; 1st strand cDNA was primed with a Not

```


	Query Match	15.9%	Score 181.6;	DB 9;	Length 524;	
	Best Local Similarity	76.4%;	Pred. No. 1.8e-28;			
	Matches 223;	Conservative	0;	Mismatches 69;	Indels 0;	Gaps 0;
QY	161	ttttcttctgttcttctgctcgtcacccaggctgaagtcacagtgcgttagctcacgg	220			
Db	124	TTTTTTTGAGATGAGTCTCGCTGTGCACCCAGCGTCGGAGTGCAGTGTGTGATCCTCG	183			
QY	221	ctcaactgcagcttgaccctcccaggctcaagtgatcctctcgctcagctttccaagtaa	280			
Db	184	CTCACTGCAACCCTCCGGTCCC GG GTTCAA GTGCTTCTCCGCTTCAGCCTCCCAAGTAG	243			
QY	281	ctgggaccacaggcatgatcaccacagctagctattgttttacatttttgttagagatg	340			
Db	244	CTGGGACTATTAGGCATCGGCCACCAAGCCTGGCTATTTTGTGTATTTTAGTAGAGACG	303			
QY	341	gggtctcaccatgttgccccagggttggtctcaaacctctgggctcaagaacatccgtcacg	400			
Db	304	GGGTTTCACCATGTGGCCAGCGTGGTCTCAAAATTCCTGACCTCAAGTGAATCCACCCACC	363			
QY	401	tcaacctccccaaatgctgggattacaaggctgaagccacccggcagggtcg	452			
Db	364	TGGGCTTCACAAAGTGTGGATTACAGGCTGAGCGCTGAGCCACTGGGCTGGCCAG	415			
RESULT 13						
AAL31088						
LOCUS						
DEFINITION						
z016c04.sl Stratagene colon (#937204) Homo sapiens cDNA clone						
IMAGE:587046 3' similar to contains Alu repetitive element; mRNA						
sequence.						
ACCESSION						
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.						
1 (bases 1 to 571)						
Hillier L., Lennon G., Becker M., Bonaldo M.F., Chiapelli B.,						
Chissoe S., Dietrich N., DuBuque T., Favello A., Gish W., Hawkins						
M., Hultman M., Kucaba T., Lacy M., Le M., Mardis E., Moore						
B., Morris M., Parsons J., Prange C., Rifkin L., Rohlfing T.,						
Schellenberg K., Soares M.B., Tan F., Thierry-Mieg J., Trevaskis E.,						
Underwood K., Wohlmann P., Waterston R., Wilson R. and Marra M.						
Generation and analysis of 280,000 human expressed sequence tags						
Genome Res. 6 (9), 807-828 (1996)						
97044478						
Contact: Wilson RK						
Washington University School of Medicine						
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108						
Tel: 314 286 1800						
Fax: 314 286 1810						
Email: est@watson.wustl.edu						
This clone is available royalty-free through LLNL : contact the						
IMAGE Consortium (info@image.llnl.gov) for further information.						
Insert Length: 2165 Std Error: 0.00						
Seq primer: -40x13 fwd. from Amersham						
High quality sequence stop: 194.						
FEATURES						
source						
1..571						
/organism="Homo sapiens"						
/db_xref="GDB:4619348"						
/db_xref="taxon:9606"						
/clone="IMAGE:587046"						
/clone_lib="Stratagene colon (#937204)"						
/tissue_type="tumor"						
/cell_line="T84 carcinoma cell line"						
/lab_host="SOIR cells (kanamycin resistant)"						
/note="Organ: colon; vector: pBluescript SK-; site_1:						
ECORI; Site 2: XhoI; Cloned unidirectionally. Primer:						


```

High quality sequence start: 13
High quality sequence stop: 660.
Location/Qualifiers
1. 661
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HR0597"
/dev_stage="Adult"
/note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTS PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
190 a 145 c 154 g 172 t

```

[illegible]

Search completed: September 24, 2002, 14:45:10
Job time: 16645 sec

THIS PAGE BLANK

THIS PAGE BLANK (USPTO)